AWARD NUMBER: W81XWH-14-1-0152

TITLE: Scaffold Attachment Factor B1: A Novel Chromatin Regulator of Prostate Cancer Metabolism

PRINCIPAL INVESTIGATOR: SUNGYONG YOU PhD

CONTRACTING ORGANIZATION: Cedars-Sinai Medical Center
Los Angeles, CA 90048

REPORT DATE: October 2016

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
October 2016	Final	1 Aug 2014 – 31 Jul 2016

4. TITLE AND SUBTITLE

Scaffold Attachment Factor B1: A Novel Chromatin Regulator of Prostate Cancer Metabolism

5a. CONTRACT NUMBER
W81XWH-14-1-0152

5b. GRANT NUMBER
GRANT11482230

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Sungyong You, Jayoung Kim, Michael R Freeman.

E-Mail:Sungyong.You@cshs.org

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

CEDARS-SINAI MEDICAL CENTER
8700 BEVERLY BLVD
LOS ANGELES CA 90048-1804

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

This study provides novel links between the SAFB1, AR, EZH2, and ONECUT2 and genes that regulate sterol metabolism in castration resistant prostate cancer (CRPC). Our findings to date have led to the working hypothesis that SAFB1 down-regulation promotes a phenotype in CRPC that results in conservation of residual androgen in the tumor, thereby promoting an "intracrine" mechanism of AR activation. Interestingly, our data suggest that SAFB1 may cooperate with other proteins that act to deplete androgen from the tumor. This hypothesis is consistent with our bioinformatics analysis of thousands of RNA expression profiles from human prostate cancers that we have incorporated into this study. We found that about 1/3 of human prostate cancers, including primary tumors, exhibit an "AR activation suppressed" phenotype. We thus tested the hypothesis that SAFB1 is a critical mediator of this phenotype. Genes involved in androgen signaling were significantly altered by SAFB1 perturbation in PC cells. In addition, metabolite profiling of androgen using mass spectrometry suggests that androgen signaling can be hyper-activated even with little amount of intracrine androgen. Along with this, we characterized the functions of SAFB1/ONECUT2/AR network that can directly regulate UGT2B15 and UGT2B17 expression, which is relevant to CRPC progression. Collectively these results suggest that SAFB1/ONECUT2/AR network is a therapeutic target in CRPC.

15. SUBJECT TERMS

Systems Biology, SAFB1, Prostate Cancer, Transcriptome

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
U	U	U	UU	31	19b. TELEPHONE NUMBER *(include area code)*

Table of Contents

Page

1. INTRODUCTION

Prostate cancer (PC) is a leading cause of death from cancer and no treatment for castration-resistant metastatic disease (CRPC) substantially prolongs life. Recent studies on humans and laboratory models have provided evidence that high circulating cholesterol is a risk factor for aggressive PC[1-5]. We recently discovered that a protein, scaffold attachment factor B1 (SAFB1), is a novel regulator of the androgen receptor and other proteins associated with prostate cancer progression to end-stage disease[6]. The purpose of my research in this project is to identify and functionally characterize the gene regulatory networks controlled by SAFB1 in human PC cells.

This project is testing the hypotheses that ***(1) SAFB1 regulates a transcriptional program that leads to PC progression when perturbed by SAFB1 loss; and that (2) down-regulation of SAFB1 promotes CRPC in part through upregulation of cholesterol-dependent intracrine androgen signaling.*** To this end, we performed chromatin immunoprecipitation-next generation DNA sequencing (ChIP-seq) and integrative network modeling to identify the SAFB1 cistrome and the extent of transcriptional collaboration of SAFB1, AR, and EZH2 in PC cells. These studies have been aided by our assembly and study of a large integrated transcriptome database of PC gene expression profiles of human tumors, which we refer to as the prostate cancer transcriptome atlas (PCTA). During the second year of the funding period, we tested whether cholesterol alters intracrine androgen mechanisms in a SAFB1-dependent manner. To this end, we applied a set of experimental tools, including metabolite profiling using mass spectrometry, qRT-PCR, and ChIP-PCR coupled with bioinformatics strategies to understand the function of the SAFB1/ONECUT2/AR network in PC, and to develop approaches directed toward targeting it.

Specific Aim 1. To characterize the SAFB1 cistrome in prostate cancer cells and to determine the metabolic and biologic effects of SAFB1 loss.
Specific Aim 2. To test whether cholesterol alters intracrine androgen mechanisms in a SAFB1-dependent manner.

2. KEYWORDS

Systems Biology, SAFB1, Prostate Cancer, Transcriptome

3. ACCOMPLISHMENTS

What were the major goals of the project?

Training Goal 1: Training and educational development in prostate cancer research
- Milestone: Presentation of project data at a national meeting
- Target months: 24
- Percentage of completion: 100%

Research Goal 1: To characterize the SAFB1 cistrome in prostate cancer cells and to determine the metabolic and biologic effects of SAFB1 loss.
- Milestones:
  1) Characterization of the SAFB1 cistrome in the presence- or absence of dihydrotestosterone (DHT).
  2) Determination of the overlapping target genes or sub-network between SAFB1 and AR or EZH2 and the genes or pathways involved in sterol metabolism and chromatin regulation.
  3) Determination of the genes or pathways strongly associated with SAFB1 regulation and PC progression.
- Target months: 12
- Percentage of completion: 100%

Research Goal 2: To test whether cholesterol alters intracrine androgen mechanisms in a SAFB1-dependent manner.
- Milestones:

1) Identification of critical regulatory nodes in the androgen metabolism network.
2) Characterization of the involvement of SAFB1 regulation of the UGT2B gene family, androgen metabolism, and downstream effects relevant to disease progression.
Target months: 24
Percentage of completion: 100%

What was accomplished under these goals?

These studies have identified novel links between the SAFB1, AR, EZH2, and ONECUT2 genes in the regulation of sterol metabolism in CRPC. Our findings to date have led to the working hypothesis that SAFB1 down-regulation promotes a phenotype in CRPC that results in conservation of residual androgen in the tumor, thereby promoting an "intracrine" mechanism of AR activation. In contrast, SAFB1 appears to cooperate with EZH2 in silencing genes in a manner that results in a manner that opposes the AR activation signature that reflects the conventionally understood pattern AR transcriptional activity. Interestingly, our data suggest that SAFB1 may cooperate with other proteins that act to ***deplete*** androgen from the tumor. This hypothesis is consistent with our bioinformatics analysis of thousands of RNA expression profiles from human PCs that we have incorporated into this study. We found that about 1/3 of human PCs, including primary tumors, exhibit an "AR activation suppressed" phenotype. We thus tested the hypothesis that SAFB1 is a critical mediator of this phenotype. Genes involved in androgen signaling were significantly altered by SAFB1 perturbation in PC cells. In addition, metabolite profiling of androgen using mass spectrometry suggests that androgen signaling can be hyper-activated even with very low amounts of intracrine androgen. Along with this, we characterized the functions of the SAFB1/ONECUT2/AR network that we have shown can directly regulate UGT2B15 and UGT2B17 expression, demonstrating relevance to CRPC progression. Collectively these results suggest that the SAFB1/ONECUT2/AR network is a therapeutic target in CRPC.

Major accomplishments include:

1) We identified the chromatin binding sites by SAFB1 by global analysis.

Identifying chromatin sites bound by SAFB1 in prostate cancer cells using chromatin immunoprecipitation and next generation DNA sequencing (ChIP-seq): Due to the limited binding affinity of SAFB1 antibody (Sigma-Aldrich), endogenous SAFB1 binding DNA fragments could not be enriched for ChIP-seq analysis. Thus, in order to increase precipitation efficacy on SAFB1, a SAFB1 expressing vector construct with an HA tag was transfected into LNCaP cells and precipitated with HA tag antibody for construction of a ChIP-seq library. To characterize the SAFB1 cistrome, chromatin sites bound by SAFB1-HA were identified using ChIP-seq. LNCaP cells were treated with 1 nM DHT or vehicle and chromatin immunoprecipitation was performed with HA tag antibody at 4 hour time points using an optimized ChIP protocol. ChIP DNA was converted into libraries and was sequenced using the Illumina HiSeq2000.

Conducting computational analysis of ChIP-seq data for SAFB1 cistrome: For identification of chromatin binding sites of SAFB1, sequencing data was processed using the Illumina analysis pipeline, aligned to the UCSC hg19/NCBI 37 version of the human genome using Bowtie[7], reads with the exact same mapping location were considered to be PCR duplicates and collapsed into a single record using samtools[8], and SAFB1-enriched binding sites were identified using the R csaw package[9]. 17,884 genome-wide SAFB1 binding sites on promoter regions (upstream 2,000 bp and downstream 500 bp from transcription start sites (TSS)) were identified by comparing with binding sites in input control sample. Overlap and feature annotation of ChIP-seq enriched regions were performed using R detailRanges function from csaw package[9]. Intersecting with 922 differentially expressed genes (DEGs) by SAFB1 knockdown, I found that 259 DEGs contain SAFB1 binding sites in their promoter regions. This result suggests that about 28% of DEGs can be regulated by SAFB1 binding in their proximal promoters. Notably, steroid and androgen metabolism related genes (ASMTL, CYP21A2, UGT2B15, UGT2B17, and HSD17B8***)*** were identified. This result is highly consistent with our preliminary data showing massive down-regulation of sterol metabolism genes with SAFB1 silencing.

2) We determined the effects of SAFB1 knockdown on the AR and EZH2 cistromes.

Performing ChIP-seq using anti-specific antibodies against AR or EZH2: ChIP-seq analysis was performed to identify AR and EZH2 target genes dependent on SAFB1 loss in LNCaP cells in the presence or absence of DHT. LNCaP SAFB1 knockdown and control cells were treated with 1 nM DHT and chromatin immunoprecipitation was performed with AR and EZH2 antibody at 4 hour time points. ChIP DNA was converted into libraries and was sequenced using the Illumina HiSeq2000.

Conducting computational analysis of ChIP-seq data for the AR and EZH2 cistromes: ChIP-seq reads were mapped to the UCSC hg19/NCBI 37 version of the human genome using Bowtie[7]. Differential AR binding sites between the SAFB1 knockdown LNCaP cells and the control cells were found by using R csaw package[9]. As an additional filter, low-abundance windows contain no binding sites were filtered out. This improves power by 1) removing irrelevant tests prior to the multiple testing correction; 2) avoiding problems with discreteness in downstream statistical methods; and 3) reducing computational work for further analyses[10]. Filtering is performed using the average abundance of each window. Binding sites are only retained if they have abundances 10-fold higher than the background. This removes a large number of binding sites that are weakly or not marked and are likely to be irrelevant.

In order to compare the list of genes associated with AR and EZH2 binding peaks to the list of genes differentially expressed on SAFB1 knockdown, the list of gene symbols for promoters associated with AR or EZH2 binding sites were intersected with the list of gene symbols for DEGs. As a result, the AR and EZH2 ChIP-seq data in SAFB1 knockdown LNCaP cells produced 7,193 and 8,038 sites compared to control cells, respectively. The 922 DEGs in SAFB1 knockdown were intersected with those genes identified as having a proximal or nearby AR and/or EZH2 binding sites in the either knockdown cells or control cells. PSMB8, HLA-B, UGT2B10, UGT2B15, KLK3 (PSA), IRS1, ABCF1, FDPS, ONECUT2, and SAFB1 were identified with significant increased or decreased binding (>1.5 fold) of AR in their promoter regions (Figure 1). This data suggests that there is a close relationship between SAFB1, AR, and EZH2 binding, and gene expression; on SAFB1 knockdown, genes associated with an AR binding peak are significantly more likely to be differentially expressed than other genes.

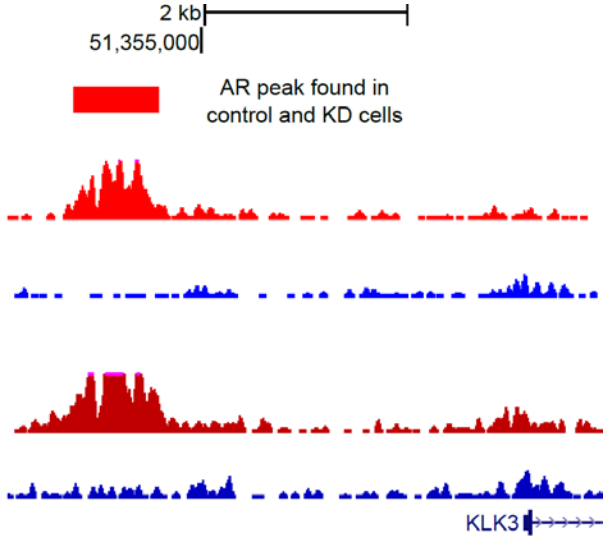


Figure 1. AR peaks on proximal promoter of KLK3 gene also known as prostate specific antigen (PSA) in both control and SAFB1 knockdown cells.

Validation of genes with AR and/or EZH2 binding sites identified in SAFB1 knockdown cells: LNCaP cells with the stable knockdown of SAFB1 using shRNA from Sigma Aldrich. The cells were analyzed for SAFB1 loss and amplified for AR protein expression and AR transcriptional activity[6].

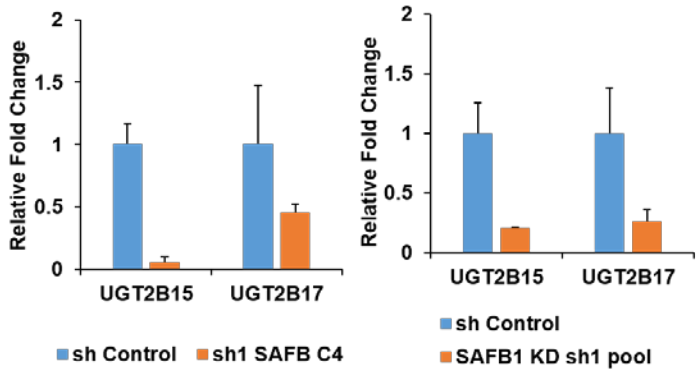


Figure 2. Decrease of UGT2B15 and 17 gene expression by SAFB KD in LNCaP (left) and 22Rv1 (right).

These SAFB1 knockdown cells showed downregulation of several members of the UGT2B family of genes, including UGT2B15 and UGT2B17 (Figure 2), the most well studied UGT2B genes within the prostate. These results were validated by qPCR using primers generated by Ohno et al.[11] and shown to be specific for the different family members. Applied Biosystems ABI Prism 7900HT qPCR machine was used to perform the analysis.

UGT2B15 and 17 gene expression changes by SAFB1 overexpression (OE) were analyzed in LNCaP cells. qPCR analysis of UGT2B15 and 17 in LNCaP cells was done after transient OE of SAFB1-HA tag (pBABE vector backbone) for 48 hours using Lipofectamine LTX (Invitrogen). Overexpression of SAFB1 was confirmed by qPCR (Figure 3A). Then, qPCR analysis was performed to measure UGT2B15 and 17 gene expression changes by SAFB1 OE. The qPCR primers and the qPCR equipment are the same as above. Significant increase of UGT2B15 and UGT2B17 gene expression were confirmed (Figure 3B). To validate whether this expression changes are directed by SAFB1 binding in promoter regions of UGT2B15 and UGT2B17 genes, we performed luciferase analysis of UGT2B15 and UGT2B17 activity in 22Rv1 PC cells (Figure 3C). For this analysis, the UGT2B15 and 17 luciferase promoter constructs (PGL4.10 vector backbone from promega) were co-transfected with control or SAFB1-HA overexpression vector into 22RV1. Baseline activity was generated from empty luciferase PGL4.10, data was normalized to this negative control (set at 100% activity). The result shows that significant increase of promoter activity of UGT2B15 and UGT2B17 genes by SAFB OE. This result demostrates that UGT2B15 and UGT2B17 genes are regulated by binding of SAFB1 in the promoter region.

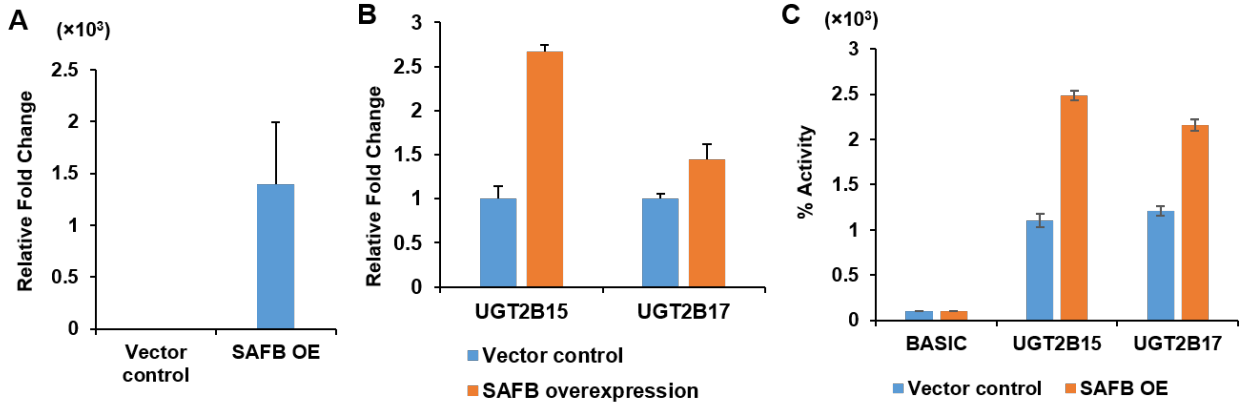


Figure 3. UGT2B15 and 17 gene expression changes and promoter activities perturbed by SAFB1 overexpression (OE) in LNCaP and 22Rv1. (A) Overexpression of SAFB1 in LNCaP. (B) qPCR analysis of UGT2B15 and 17 genes perturbed by SAFB1 overexpression in LNCaP. (C) Luciferase activity of UGT2B15 and 17 gene promoter by SAFB1 in 22Rv1.

Identifying consensus binding motifs of SAFB1 and AR: To identify a consensus binding motif for SAFB1 and AR, motif analysis was done for SAFB1 and AR ChIP-seq data sets. I found the AR/PR motif (Figure 4A) in ~60% of peaks and the AR half-site motif (Figure 4B) in ~75% of peaks. Figure 4 shows motif logos of SAFB1 and AR bining motifs from MEME analysis[12]. This result obtained from 1,069 common sitess between the SAFB1 and AR data sets.

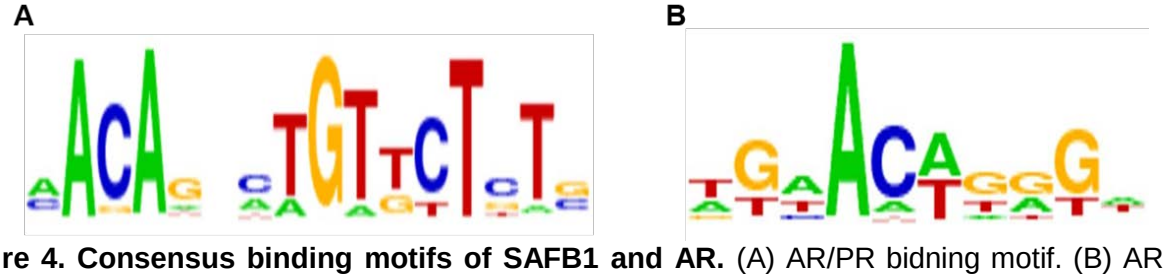
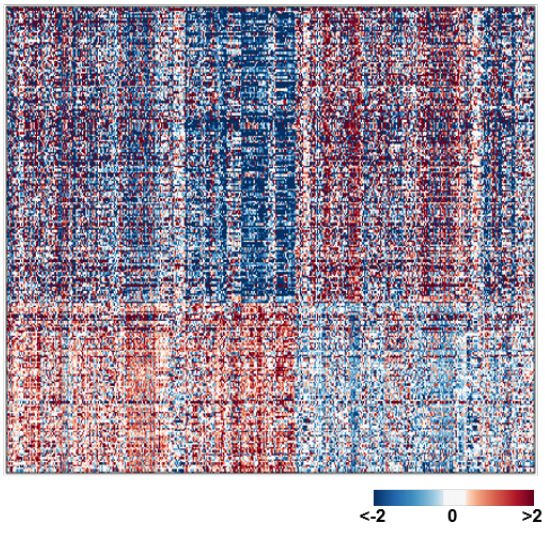


Figure 4. Consensus binding motifs of SAFB1 and AR. (A) AR/PR bidning motif. (B) AR half-site motif.

3) We found a clinical correlation of SAFB1 loss and PC progression and patient outcomes.

Transcriptome analysis revealed SAFB1 loss-dependent genes and pathways in clinical specimens: In order to identify genes in human PC tumors that correlate with alterations in SAFB1 gene expression, I compared 726 prostate tumor samples with low (<25 percentile) vs. high (>75 percentile) expression of SAFB1. Over 3,000 differentially expressed genes (DEGs) between prostate tumors with low (or no) and high expression of SAFB1 were selected with false dicovery rate (FDR)<0.05, and applied to functional enrichment analysis using DAVID software (Figure 5A and B). Enriched cellular processes indicate that SAFB1-dependent differential expression results in a more aggressive phenotype, including increases in steroid hormone reponses, regulation of blood vessel formation (angiogenesis), and regulation of RNA processing (Figure 5B). By integrating SAFB1 and AR ChIP-seq data sets and differential expression of SAFB1 knockdown cells with differential expression of prostate cancer patients, 387 genes were identified. These genes are differentially expressed in both SAFB1 knockdown cells and prostate cancer patients with low SAFB1 expression, as well as have SAFB1 and/or AR binding in their promoters. Among these genes, PSMB8, HLA-B, UGT2B10, UGT2B15, KLK3 (PSA), IRS1, ABCF1, FDPS, ONECUT2, and SAFB1 were also identified.

A



B

Class	Term	P Value
Up-regulated in low SAFB	response to steroid hormone stimulus	3.42E-06
	regulation of tube size	1.38E-05
	regulation of blood vessel size	1.38E-05
	circulatory system process	1.69E-05
	blood circulation	1.69E-05
Down-regulated in low SAFB	RNA processing	1.09E-13
	mRNA processing	6.55E-13
	mRNA metabolic process	6.55E-13
	RNA splicing	1.09E-12
	RNA splicing, via transesterification reactions	2.16E-09

Figure 5. Differentially expressed gene by SAFB1 loss in clinical samples and their enriched cellular processes. (A) Heatmap depicts differential expression pattern of SAFB1 dependent gene signnature in prostate cancer. (B) Enriched cellular processes by up- or down-regulated genes between patients with SAFB1-high (>75 percentile) and SAFB1-low (<25 percentile).

4) SAFB1 knockdown results in activation of an intracrine AR network arising from increased levels of intracellular androgen.

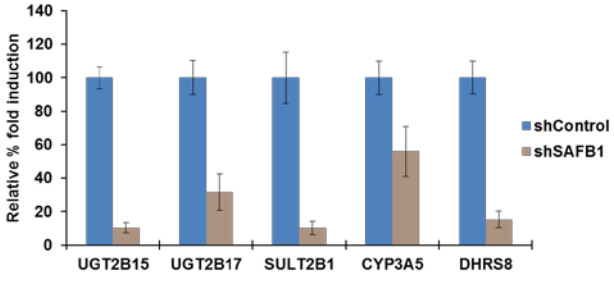


Figure 6. SAFB1 silencing in LNCaP cells downregulates genes involved in androgen catabolism

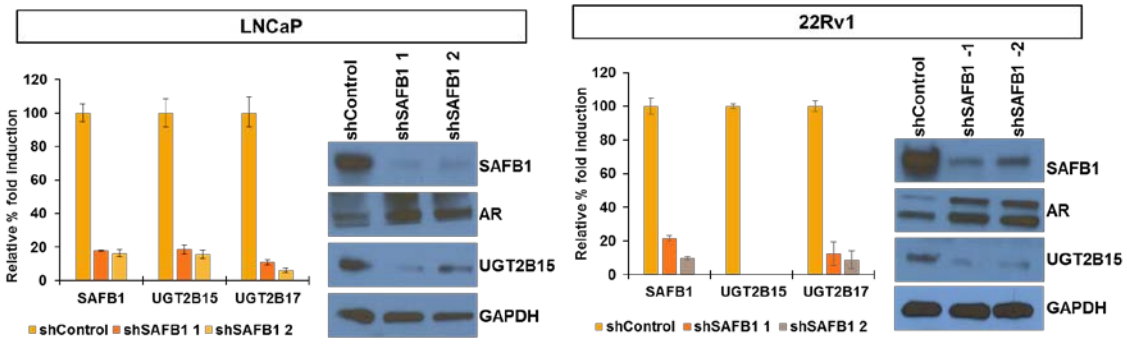
The effects of SAFB1 silencing suggest stimulation of an intracrine androgen pathway: we performed a global analysis to look at the pathways that are disrupted when SAFB1 was lost in cell line model and patient tissues. We saw that androgen metabolism was a major pathway affected by SAFB1 loss. We found that many of the genes that are significantly altered lie on the boundaries of the androgen synthesis pathways. They mainly were on the pathways that inactivate androgen. When we measure the expression of these targets that indicated in the microarray, we confirm that there is a downregulation in LNCaP-

SAFB1 KD cells (Figure 6).

This confirms the downregulation of the androgen catabolism pathways, so from this we hypothesize that the effect of SAFB1 loss in relationship to the intracrine androgen hypothesis is therefore that there is a possible increase in DHT stability. The SULT2B1 sulfotransferase is known to utilize 3-phospho-5-adenylyl sulfate as sulfonate donor to catalyze the sulfate conjugation of many hormones, neurotransmitters, drugs and xenobiotic compounds. Sulfonation increases the water solubility of most compounds, and therefore their renal excretion, but it can result in activation to form active metabolites. Sulfates hydroxysteroids like DHEA. Isoform 1 preferentially sulfonates cholesterol, and isoform 2 avidly sulfonates pregnenolone but not cholesterol. SULT2B1 and CYP3A5 oxidize the testosterone to form inactive 6β-hydroxyl testosterone, which cannot be used for conversion to DHT. DHRS8 converts 5-alpha-androstane-3alpha-17alpha diol to androstenedione, which is well known that 5-alpha-androstane-3alpha-17alpha diol is a precursor for DHT.

Figure 7. Stable SAFB1 knockdown and significant downregulation of UGT2B15 and UGT2B17 in LNCaP and 22Rv cells.

When looking at these gene sets we decided to focus on UGT2B15 and UGT2B17, which are expressed at high levels in the prostate and in the LNCaP and 22Rv1 cell line models. UGT2B15 and UGT2B17 are the principal enzymes that mediate inactivation and removal of DHT from the prostate. This process is irreversible; consequently, modulation of these genes can have potentially very significant functional consequences in prostate cancer cells. In order to study the effects of SAFB1 loss we regenerated SAFB1 KD cell lines in LNCaP and this time in another AR positive cell line 22Rv1. Here we see that the downregulation of SAFB1 using two independent hairpins can downregulate UGT2B15 and UGT2B17 expression (Figure 7). These results were confirmed by western blot. When we analyze the effect of SAFB1 KD in 22Rv1, we see the same results as in LNCaP. We have therefore seen in two independent cell lines using two different hairpins for SAFB1 KD that there is amplification of AR as well as the downregulation of UGT2B15 and UGT2B17. AR is upregulated and we could show with a published antibody for UGT2B15 that the expression of the protein is



downregulated as well.

Metabolic profiling of free DHT using mass spectrometry: To understand the consequence of downregulation of UGT2B15/17 on DHT levels, we measured free DHT levels in the media of LNCaP cells after supplementation of the media for 2, 4, 6, 8 hours with radiolabeled DHT (Figure 8). We found that SAFB1 KD cells maintain their levels of free DHT over time, while the control cells decrease in the levels of DHT over time. The stability of free DHT available for the cell to use was much higher in SAFB1 KD cells compared to control cells. These results indicate that downregulation of UGT2B15/17 alters androgen availability in a manner that is consistent with our hypothesis.

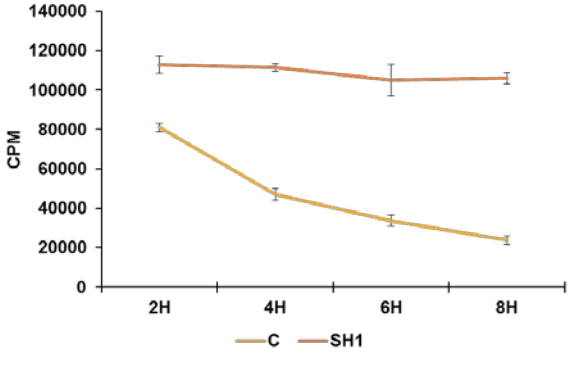


Figure 8. SAFB1 knockdown alter levels of Free DHT.

Development of a new classification scheme for prostate cancer: To better understand the molecular

heterogeneity of PC, I have assembled a “prostate cancer transcriptome atlas” (PCTA) software tool and database that contains more than 4,000 human prostate cancer transcriptomes assembled from public databases and the literature (including GEO, Array Express and TCGA). Using the PCTA, I examined transcriptome-based patterns of diverse oncogenic pathways and other important features in PC using a collection of 22 previously published gene expression signatures[13-29], resulting in a summary of activity score data of 14 pathways of the tumors. When applied an unsupervised clustering algorithm based on non-negative matrix factorization (NMF)[30] to pathway activity score data consisting of 1,321 prostate tumors, I identified three distinct sub-groups, shown below as Group 1-3[31] (Figure 9).

The heatmap in Figure 9 shows the surprising result that identifiable molecular features are evident across all disease categories through Gleason Score (GS) <7 to metastatic or castration-resistant PC (CRPC/Met), suggesting that prostate tumors retain identifiable epigenomic properties as tumor evolution proceeds. Although there are exceptions to the broad patterns, we found a remarkable consistency within groups. ERG fusion-inducible gene expression is predominant in Group 1, which is also characterized by high AR activation activity scores. AR-variant inducible gene expression is clustered in Group 2, which also shows high proliferation and neuroendocrine activity. In contrast to the features seen in Group 1 and Group 2, Group 3 is uncharacterized as a distinct entity in prostate cancer. Group 3 exhibits pro-neural and mesenchymal activation signatures. Notably, the AR activation signature and AR variant-inducible signatures are relatively low in Group 3.

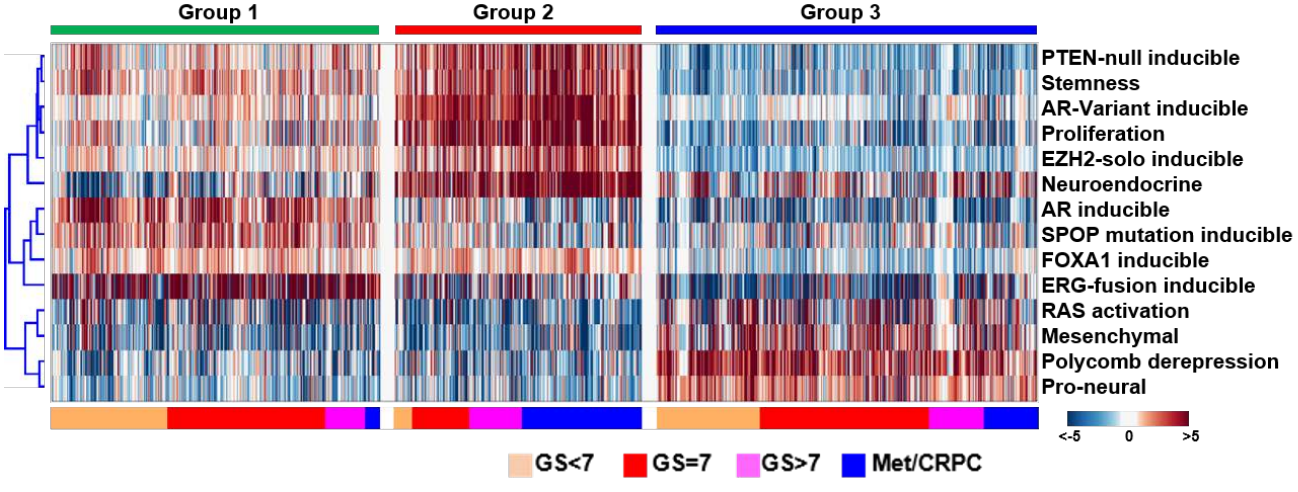


Figure 9. Patterns of signature pathway activities of 1321 prostate cancer patients in transcriptome atlas. (A) Patterns of activity scores were determined for each sample using Z score method. Consensus NMF clustering of 1321 prostate tumors using 14 pathway activity scores revealed three intrinsic molecular subtypes of prostate cancer (Group 1-3). The pathway activity scores (y-axis) were clustered by complete linkage hierarchical clustering method.

<u>Validation of the three subtypes using independent cohorts:</u> I have validated this classification system in 10 independent patient series, consisting of over 1,200 RNA expression profiles (Figure 10). This result suggests that it might be possible to cluster essentially all prostate cancers into one of only three subtypes defined by gene signatures that have been functionally implicated in the disease.

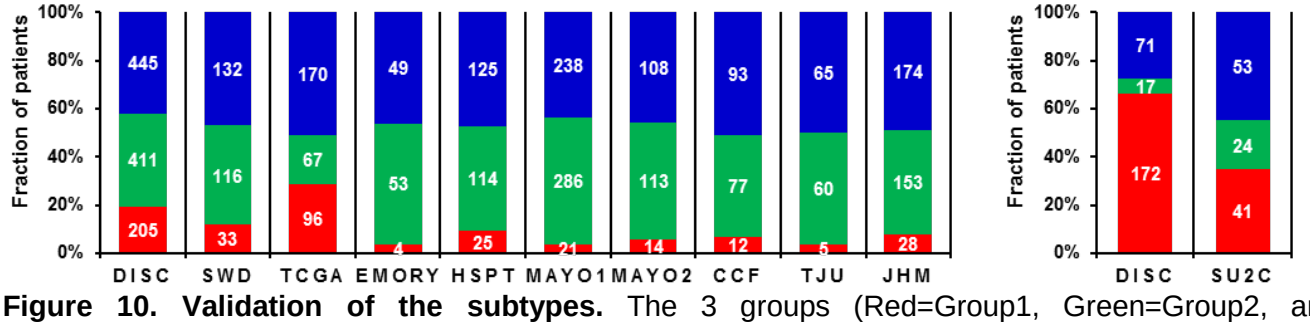


Figure 10. Validation of the subtypes. The 3 groups (Red=Group1, Green=Group2, and Blue=Group3) were recognized in 10 independent cohorts. Comparable fractions of patients with

primary prostate tumors (left) and CRPC/Met (right) are assigned to each subtype within the different cohorts. DISC=discovery cohort; SWD=Swedish watchful waiting cohort; TCGA=TCGA cohort; EMORY=Emory cohort; HSPT= Health Study Prostate Tumor cohort; MAYO1=Mayo clinic cohort 1; MAYO2=Mayo clinic cohort 2; CCF=Cleveland clinic cohort; TJU=Thomas Jefferson University cohort; SU2C= SU2C/PCF Dream Team cohort.

Discovering a novel driver of aggressive prostate cancer variants: To computationally identify transcription factors (TFs) that are highly active in this disease space, I used the large number of CRPC/Met tumors (n=260) in the PCTA. We integrated RNA expression data with TF-target gene interaction data collected from a number of chromatin immunoprecipitation (ChIP) and curated databases that contain genes that share TF binding sites. We then conducted a master regulator analysis (MRA) based on a combination of gene set enrichment analysis (GSEA) and rank correlation of TF expression level and RNA expression level of known targets for each TF. This analysis identified a set of TFs known to be functionally significant in CRPC/Met PC, including AR, EZH2, FOXM1, and E2F3, thereby validating our approach. Surprisingly, this analysis also identified a TF that has not been studied in PC, ONECUT2, an atypical homeobox TF that has been implicated in liver, pancreas and neural development. Notably, ONECUT2 is one of the SAFB1 target genes in PC cells. We found that ONECUT2 gene expression is significantly down-regulated by SAFB1 KD in LNCaP (Figure 11). ONECUT2 expression gradually increases across the disease categories in the PCTA data set, and bionformatics modeling predicts that it functionally interacts with AR, EZH2, and FOXA1. Enforced and silenced ONECUT2 in LNCaP and 22Rv1 cells were done by transfecting shONECUT2 and ONECUT2-overexpressing vector construct, and conducted oligonucleotide expression array and functional experiments. Significantly, ONECUT2 can potently inhibit AR, PSA, EZH2, and FOXA1 expression (Figure 11), consistent with our computational modeling predictions.

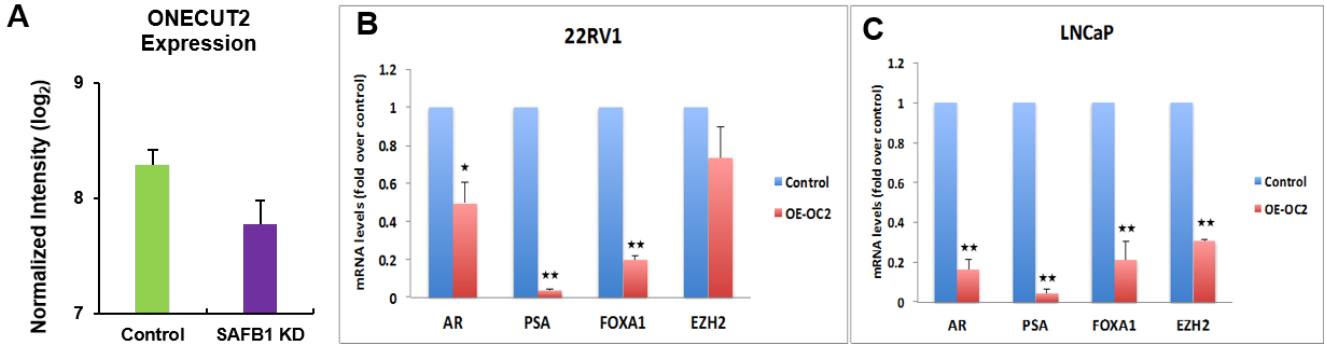


Figure 11. Gene expression of ONECUT2, AR, PSA, FOXA1 and EZH2. (A) Differential gene expression of ONECUT2 gene by SAFB1 knockdown in LNCaP (FDR<0.05). (B) ONECUT2 (OC2) suppresses AR, PSA, FOXA1, and EZH2 in prostate cancer cell lines. Total RNA was isolated from 22Rv1 and LNCaP cells overexpressing OC2 and real-time qPCR was performed using TaqMan probes for the indicated genes. Each value represents the mean±SEM of 3 independent experiments performed in triplicate. Significant differences are denoted by asterisks (*p≤0.05. **p≤0.01).

ONECUT2 plays a role in stimulating growth of 22Rv1 cells (Figure 12A) and therefore might be targeted in vivo to limit progression of CRPC. The PSA/KLK3 enhancer is a prostate regulatory element, strongly supporting the role of ONECUT2 in PC. From this result, I hypothesized that ONECUT2 is a driver of PSA-negative clones that may expand after therapy and ONECUT2 expression level may be an indicator of progression to metastasis (Figure 12B).

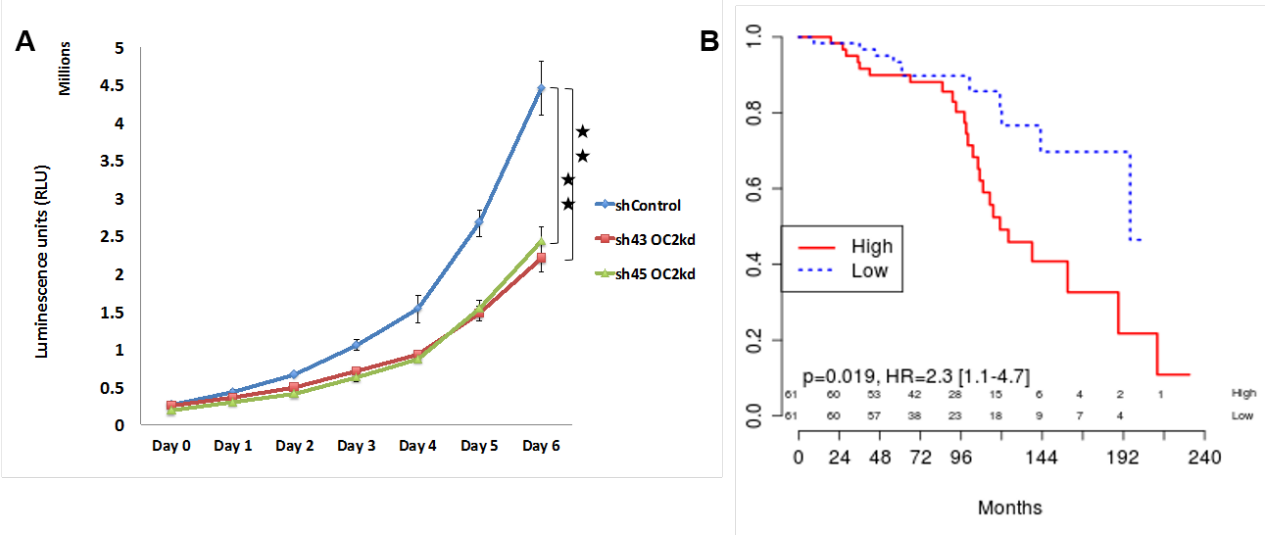


Figure 12. Biological and clinical implication of ONECUT2. (A) Proliferation assay demonstrated significant inhibition of cell proliferation by ONECUT2 knockdown (kd). (B) Kaplan-Meier analysis showing top vs. bottom tertiles of OC2 expression level in relation to metastasis-free survival (CCF cohort).

We employed the PC classification scheme that we developed in order to ask whether ONECUT2 activity segregates between Groups 1-3. We used the gene expression perturbation data generated using enforced and silencing methods to nominate ONECUT2 activation and repression signatures. We then applied these signatures to the three subtypes developed from the PCTA data. We found that the ONECUT2 activation signature is most active in Group 3 in all disease categories, but that the ONECUT2 repression signature increases progressively in Group 2, with highest activity in CRPC/Met tumors (Figure 13). These findings demonstrate that we can map master regulator activity onto human PC by integrating our classification scheme with laboratory data.

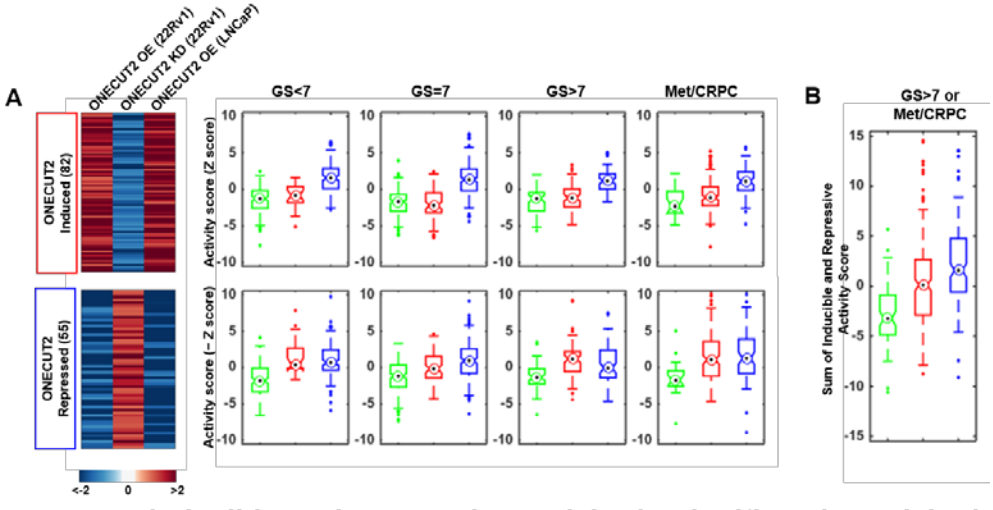


Figure 13. ONECUT2-inducible and -repressive activity is significantly enriched in group 3 in comparison to the other 2 groups. (A) Heatmaps show differential expression patterns of genes perturbed by ONECUT2 overexpression and knockdown in 22Rv1 and LNCaP cells (FDR<0.05 and fold change ≥2). Results from the PCTA cohort are shown in the panels A and B. Group 1 = green, group 2 = red, group 3 = blue.

5) An intracrine AR network responds to decreases and increases in cholesterol levels

Revealing the novel regulatory relationship between OENCUT2 and UGT2B15: We next asked, whether ONECUT2 regulates UGT2B15 and SAFB1 is required for its regulation. We thus tested the impact of enforced expression of ONECUT2 in the regulation of UGT2B15 gene expression using the

LNCaP and 22Rv1 cell line models by differentially regulating SAFB1 expression using shRNAs (Figure 14). As a result, enforced expression of ONECUT2 significantly increased the expression of UGT2B15 gene in 10 nM DHT condition compared to 0 nM DHT as shown in left panel of Figure 14. However, ONECUT2 expression was not significantly altered by knockdown of SAFB1 as shown in the middle of Figure 14. We also found that ONECUT2 overexpression in the context of knockdown of SAFB1 do not maintain UGT2B15 gene expression under 10nM DHT treatment as shown in right panel of Figure 14. This suggests that expression of UGT2B15 gene is regulated by ONECUT2 and is mediated by SAFB1 in the context of high DHT concentration.

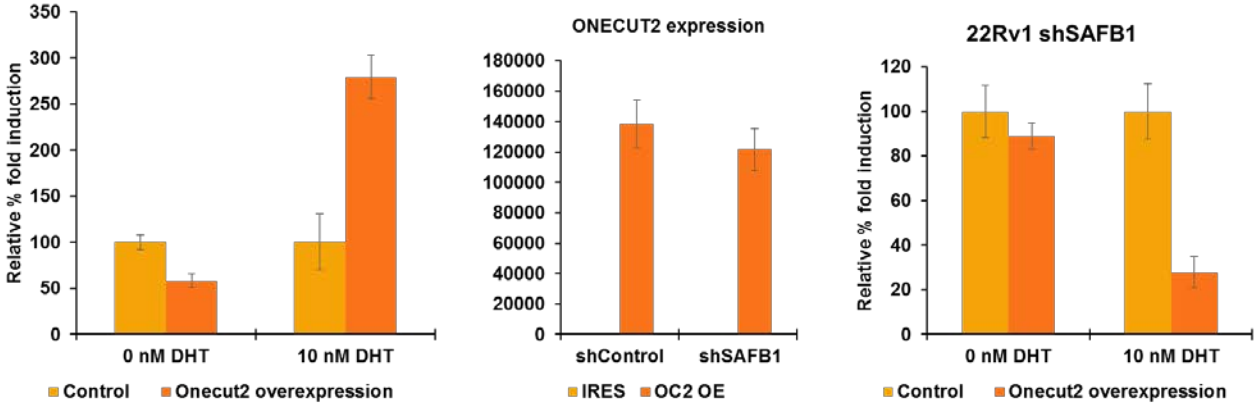


Figure 14. ONECUT2 is a positive regulator of UGT2B15.

Correlation of SAFB1/ONECUT2/AR with UGT2B15 and UGT2B17 in CRPC: We have shown that SAFB1 and ONECUT2 co-regulate UGT2B15 expression. We thus predicted that there would be a correlation with the expression of these genes in CRPC patients (Figure 15). Using 260 samples of CRPC/Met samples from the PCTA, statistically significant positive correlation between ONECUT2, SAFB1, and UGT2B15 and UGT2B17 were observed. We then test whether AR activity correlates with these gene exoression in CRPC/Met tumors. Interestingly, AR activity shows inverse correaltion of those gene expressions. AR activity was measured based on a previously published method[32]. This lead us to test whether inverse relatioship of AR activity have something to do with cell cycle and proliferation activity (CCP). To this end. we computed CCP scores of the CRPC/Met tumors and found

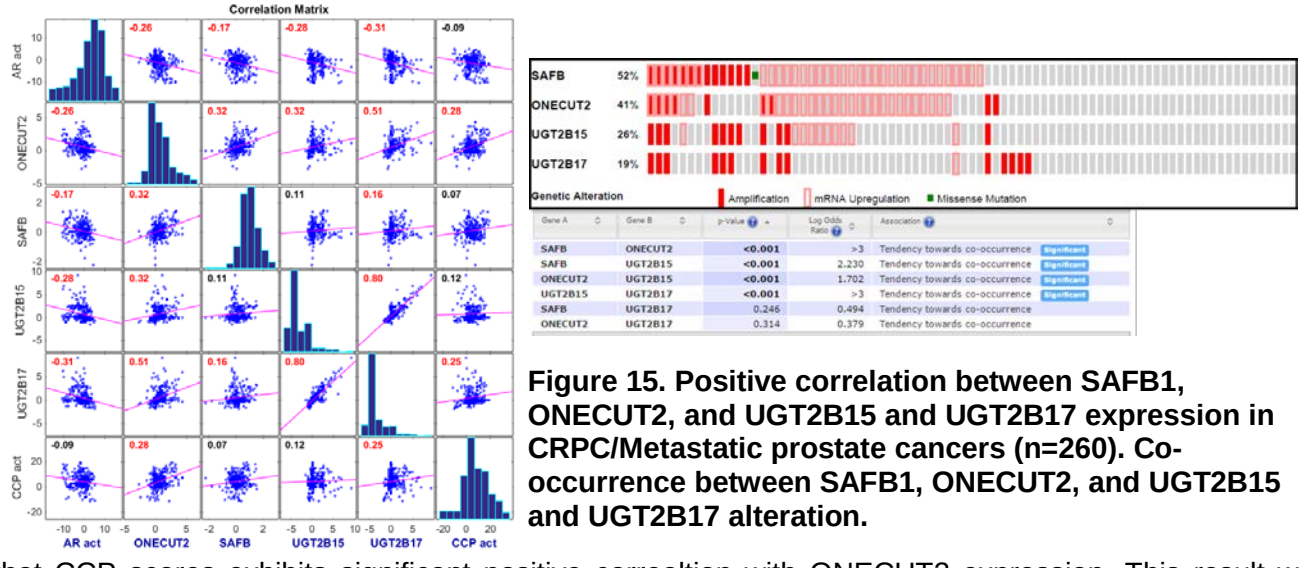


Figure 15. Positive correlation between SAFB1, ONECUT2, and UGT2B15 and UGT2B17 expression in CRPC/Metastatic prostate cancers (n=260). Co-occurrence between SAFB1, ONECUT2, and UGT2B15 and UGT2B17 alteration.

that CCP scores exhibits significant positive correaltion with ONECUT2 expression. This result was promising for us to show that SAFB1/ONECUT2 and UGT2B15 and UGT2B17 expression is correlated in patient samples.

In addition to the PCTA cases, we attepted to survey any alteration of the genes including SAFB1, ONECUT2, UGT2B15, and UGT2B17 in the patients with neuroendocrine (NE) tumors[33]. Of

note, we found significant DNA amplification and/or overexpression of the genes in NE tumors as shown in the right panel of Figure 15. Significant co-occurance of the gene alterations were evident by the co-occurance statistics in Figure 15. Given the series of data for SAFB1 network, we could modeled SAFB1 network regulation in the CRPC, which can be represented by the two regulatory directions: 1) Loss of SAFB1 can exert increase AR activity through the enhanced stability of intracrine androgen caused by the low expression of UGT2B15 and UGT2B17, which is major component of intracrine androgen catabolism, resulting in reinforced AR signaling and 2) Intact SAFB1 can drive AR independent CRPC in interaction with ONECUT2 through a downregulation of intracrine androgen by the upregulation of UGT2B15 and UGT2B17.

Characterization of the SAFB1/AR/ONECUT2/UGT2B15/UGT2B17 network: Given the SAFB1 network model, we attempted to investigate if SAFB1/ONECUT2/AR assemble the complex and positively regulates UGT2B15 (Figure 16). We thus performed a co-IP of AR and SAFB1 using the corresponding antibodies. We predicted from the model above that SAFB1, ONECUT2, and AR bind in a complex and bind to UGT2B15 promoter to regulate its expression. We already know that AR binds to the UGT2B15 and UGT2B17 promoter in a region proximal to the TSS. We used the same ChIP-qPCR primer sets to examine the SAFB1, ONECUT2 and AR binding to the UGT2B15 and UGT2B17 promoters and we saw the interactions for all three proteins from the co-IP data. Collectively, the result provides evidence that SAFB1, ONECUT2 and AR form a complex to regulate UGT2B15 and UGT2B17 gene expression.

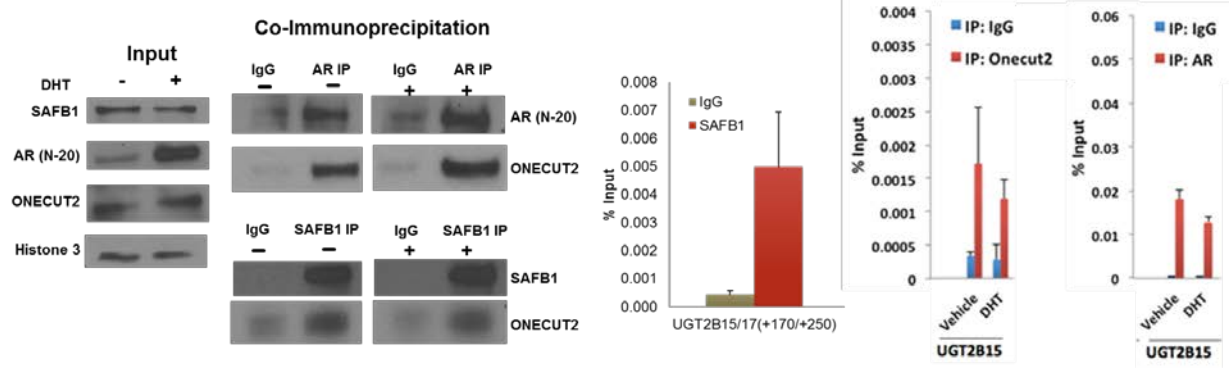


Figure 16. SAFB1, ONECUT2 and AR are interact with each other and bind to and regulates the UG2B15/17 promoters.

Therapeutic implication of SAFB1/ONECUT2/AR network: From the model validation above, we found that SAFB1 loss promotes i) hyperactive AR; ii) stable levels of active DHT; and iii) positive regulation of UGT2B15 and UGT2B17. These 3 findings probably lend to its ability to resist potent CRPC treatments used currently. We therefore measured the cell proliferation of SAFB1 KD cells and control cells in the presence of different concentrations of Enzalutamide (Figure 17). We could see that proliferation rate of the knockdown cells is higher than one of the control cells in the presence of the drug. This suggests that SAFB1 loss confers a resistance to Enzalutamide treatment. This result lead us to investigate if SAFB1 expression in the CRPC/Met patients exhibit any association with anti-androgen therapys. Using the previously published gene expression data[34], we found that patients who progressed to metastatic PC even after abiraterone (Abi) or Enzalutamide (Enz) treatment had lower levels of SAFB1 expression compared to the patients without treatment. This suggests that SAFB1 loss confers resistance to in our case at least Enzalutamide.

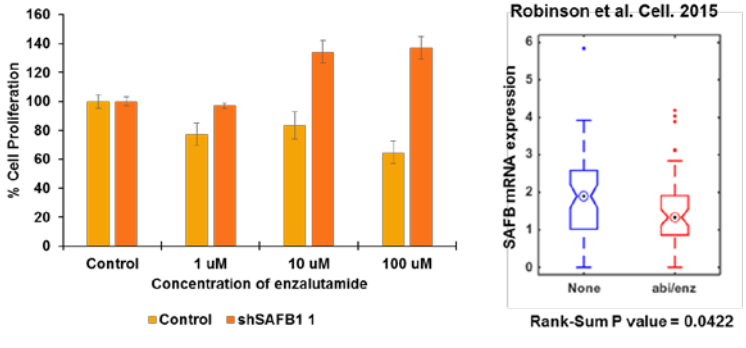


Figure 17. SAFB1 knockdown confers enzalutamide resistance.

6) Key research accomplishments

- Generation of the first ChIP-seq analysis of SAFB1 and the first identification of the SAFB1 cistrome in PC cells.
- Discovery that UGT2B15 and UGT2B17 are regulated by SAFB1, indicating that these androgen-inactivating genes are a component of the SAFB1 transcriptional network.
- Development of a novel classification system for PC that has utility in providing novel and actionable clinical information.
- Identification of ONECUT2 as a novel driver of aggressive PC variants.
- UGT2B15 and UGT2B17 genes demonstrated to reside within this AR-metabolic network
- Charactering novel interactions between SAFB1, AR, and ONECUT2 in PC
- UGT2B15 and UGT2B17 direct target of SAFB1 network in PC

7) Conclusion

ChIP-seq analysis followed by computational analysis permitted the determination of the extent to which chromatin occupancy of SAFB1 cistrome components reflects gene expression patterns characteristic of AR and EZH2 activity. Integration of our own ChIP-seq data and patient gene expression profiles allowed us to identify the extent of transcriptional collaboration of SAFB1, AR, and ONECUT2 in PC cells and human prostate tumors. UGT2B15 and UGT2B17 expression were coordinately regulated in many aggressive CRPC/Met patient samples through the interactions of SAFB1, AR, and ONECUT2 in a context of distinct cholesterol levels. Collectively, these data define a novel type of CRPC that does not function by AR hyperactivation, and which may be independent of intracellular androgen and AR activity.

8) Other achievement

We found that UGT2B15 and UGT2B17 gene expression is significantly increased in androgen independent LNCaP clones (C-81). Using this system, we have developed a platform to measure metabolic changes (such as DHT) by modulation of metabolic genes regulated by SAFB1. (Figure 18)

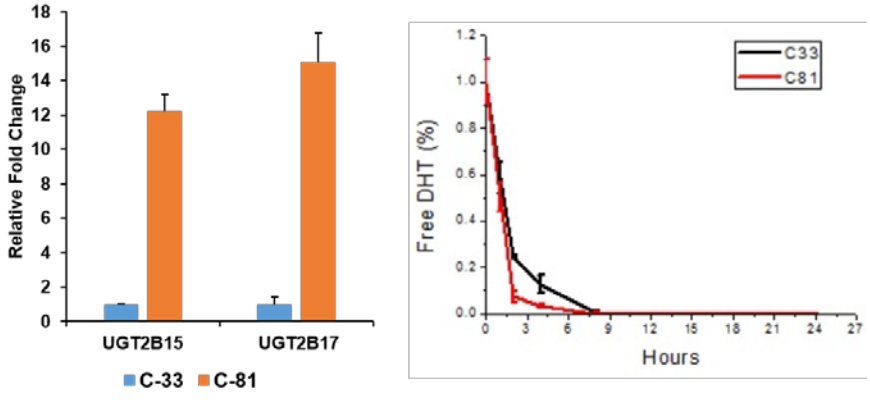


Figure 18. Development of measure of metabolite changes. (Left) LNCaP derived C33 and C81 cells generated by the lab of Min Fong Lin are a cell line that were passaged in 5% fetal bovine serum for 33 passage (C33) and 81 passages (C81). The C81 cell line is hormone insensitive while C33 cells are hormone sensitive. Upon comparing these cells, there is a large increase ~10-15 fold) in UGT2B15 & 17 in the hormone insensitive cell line C81 in comparison to hormone sensitive C33. (n=3). (Right) LNCaP derived C33 and C81 cells were analyzed by HPLC in the lab of Nima Shariffi. C81 has a (slightly) stronger activity as free DHT signal decreased faster at 2 hours. The method employed was to treat 1 million cells with 100nM cold plus some hot DHT, and examine hydrophobic radioactive signals (majorly DHT) in culture media by HPLC. (n=3)

9) Reference

1. Solomon KR, Freeman MR. Do the cholesterol-lowering properties of statins affect cancer risk? Trends Endocrinol Metab. 2008;19: 113-121.

2. Platz EA, Leitzmann MF, Visvanathan K, et al. Statin drugs and risk of advanced prostate cancer. J Natl Cancer Inst. 2006;98: 1819-1825.
3. Flick ED, Habel LA, Chan KA, et al. Statin use and risk of prostate cancer in the California Men's Health Study cohort. Cancer Epidemiol Biomarkers Prev. 2007;16: 2218-2225.
4. Murtola TJ, Visakorpi T, Lahtela J, Syvala H, Tammela T. Statins and prostate cancer prevention: where we are now, and future directions. Nat Clin Pract Urol. 2008;5: 376-387.
5. Hamilton RJ, Banez LL, Aronson WJ, et al. Statin medication use and the risk of biochemical recurrence after radical prostatectomy: results from the Shared Equal Access Regional Cancer Hospital (SEARCH) Database. Cancer. 2010;116: 3389-3398.
6. Mukhopadhyay NK, Kim J, You S, et al. Scaffold attachment factor B1 regulates the androgen receptor in concert with the growth inhibitory kinase MST1 and the methyltransferase EZH2. Oncogene. 2013.
7. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009;10: R25.
8. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25: 2078-2079.
9. Lun AT, Smyth GK. De novo detection of differentially bound regions for ChIP-seq data using peaks and windows: controlling error rates correctly. Nucleic Acids Res. 2014;42: e95.
10. McCarthy DJ, Chen Y, Smyth GK. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. Nucleic Acids Res. 2012;40: 4288-4297.
11. Ohno S, Nakajin S. Determination of mRNA expression of human UDP-glucuronosyltransferases and application for localization in various human tissues by real-time reverse transcriptase-polymerase chain reaction. Drug Metab Dispos. 2009;37: 32-40.
12. Bailey TL, Johnson J, Grant CE, Noble WS. The MEME Suite. Nucleic Acids Res. 2015;43: W39-49.
13. Phillips HS, Kharbanda S, Chen R, et al. Molecular subclasses of high-grade glioma predict prognosis, delineate a pattern of disease progression, and resemble stages in neurogenesis. Cancer Cell. 2006;9: 157-173.
14. Beltran H, Rickman DS, Park K, et al. Molecular characterization of neuroendocrine prostate cancer and identification of new drug targets. Cancer Discov. 2011;1: 487-495.
15. Assou S, Le Carrour T, Tondeur S, et al. A meta-analysis of human embryonic stem cells transcriptome integrated into a web-based expression atlas. Stem Cells. 2007;25: 961-973.
16. Yu J, Yu J, Rhodes DR, et al. A polycomb repression signature in metastatic prostate cancer predicts cancer outcome. Cancer Res. 2007;67: 10657-10663.
17. Bild AH, Yao G, Chang JT, et al. Oncogenic pathway signatures in human cancers as a guide to targeted therapies. Nature. 2006;439: 353-357.
18. Lee TI, Jenner RG, Boyer LA, et al. Control of developmental regulators by Polycomb in human embryonic stem cells. Cell. 2006;125: 301-313.
19. Xu K, Wu ZJ, Groner AC, et al. EZH2 oncogenic activity in castration-resistant prostate cancer cells is Polycomb-independent. Science. 2012;338: 1465-1469.
20. Geng C, Rajapakshe K, Shah SS, et al. Androgen receptor is the key transcriptional mediator of the tumor suppressor SPOP in prostate cancer. Cancer Res. 2014;74: 5631-5643.
21. Robinson JL, Hickey TE, Warren AY, et al. Elevated levels of FOXA1 facilitate androgen receptor chromatin binding resulting in a CRPC-like phenotype. Oncogene. 2014;33: 5666-5674.
22. Jin HJ, Zhao JC, Ogden I, Bergan RC, Yu J. Androgen receptor-independent function of FoxA1 in prostate cancer metastasis. Cancer Res. 2013;73: 3725-3736.
23. Iljin K, Wolf M, Edgren H, et al. TMPRSS2 fusions with oncogenic ETS factors in prostate cancer involve unbalanced genomic rearrangements and are associated with HDAC1 and epigenetic reprogramming. Cancer Res. 2006;66: 10242-10246.
24. Setlur SR, Mertz KD, Hoshida Y, et al. Estrogen-dependent signaling in a molecularly distinct subclass of aggressive prostate cancer. J Natl Cancer Inst. 2008;100: 815-825.
25. Nelson PS, Clegg N, Arnold H, et al. The program of androgen-responsive genes in neoplastic prostate epithelium. Proc Natl Acad Sci U S A. 2002;99: 11890-11895.
26. Sharma NL, Massie CE, Ramos-Montoya A, et al. The androgen receptor induces a distinct transcriptional program in castration-resistant prostate cancer in man. Cancer Cell. 2013;23: 35-47.
27. Mendiratta P, Mostaghel E, Guinney J, et al. Genomic strategy for targeting therapy in castration-resistant prostate cancer. J Clin Oncol. 2009;27: 2022-2029.

28. Tomlins SA, Laxman B, Varambally S, et al. Role of the TMPRSS2-ERG gene fusion in prostate cancer. Neoplasia. 2008;10: 177-188.
29. Hu R, Lu C, Mostaghel EA, et al. Distinct transcriptional programs mediated by the ligand-dependent full-length androgen receptor and its splice variants in castration-resistant prostate cancer. Cancer Res. 2012;72: 3457-3462.
30. Brunet JP, Tamayo P, Golub TR, Mesirov JP. Metagenes and molecular pattern discovery using matrix factorization. Proc Natl Acad Sci U S A. 2004;101: 4164-4169.
31. You S, Knudsen BS, Erho N, et al. Integrated Classification of Prostate Cancer Reveals a Novel Luminal Subtype with Poor Outcome. Cancer Res. 2016;76: 4948-4958.
32. Cristescu R, Lee J, Nebozhyn M, et al. Molecular analysis of gastric cancer identifies subtypes associated with distinct clinical outcomes. Nat Med. 2015;21: 449-456.
33. Beltran H, Prandi D, Mosquera JM, et al. Divergent clonal evolution of castration-resistant neuroendocrine prostate cancer. Nat Med. 2016;22: 298-305.
34. Robinson D, Van Allen EM, Wu YM, et al. Integrative clinical genomics of advanced prostate cancer. Cell. 2015;161: 1215-1228.

What opportunities for training and professional development has the project provided?

I was promoted to the rank of Instructor, and recently Assistant Professor at Cedars-Sinai, a position from which I can submit independent grant proposals. I have published a first author study in ***Cancer Research*** describing the new prostate cancer classification scheme we have developed and its possible clinical significance, which was DIRECTLY derived from this proposed study. This work has been highlighted in the Research Highlights Section of Nature Reviews Urology, under the heading “Prostate cancer: Novel subtyping could aid stratification and therapy” 2016 July 5. doi:10.1038/nrurol.2016.130 (see PRODUCTS). I also gave a poster presentation at the 2014 Annual Conference of the American Urological Association (AUA), the 2014 Prostate Cancer Foundation (PCF) meeting, the 2015 American Association for Cancer Research (AACR) Special Conference and the 2015 The Prostate Cancer Foundation (PCF) 22nd Annual Scientific Retreat (see PRODUCTS). To support education and teaching of bioinformatics and computational methods within the Cedars-Sinai prostate cancer research community, I gave presentations in lab meetings, journal club, and workgroup meetings. I have substantive one-to-one discussions with the mentors several times per week and is in near-constant contact. I have (and will continue) close communication with other senior investigators through many other routes, including (1) weekly joint lab meetings, (2) bi-weekly Cancer Biology Journal Club (organized by Dr. Kim), and (3) bi-weekly Cancer Genomics Journal Club (organized by Dr. Kim). This is a very interactive community with open lines of communication across eight nationally prominent prostate cancer research laboratories, where opinions, regents and data are continuously shared.

How were the results disseminated to communities of interest?

I created a large (>4,000 specimens) RNA expression data set from prostate cancer and benign prostate tissue. From this large data set, he has demonstrated for the first time that prostate cancers, possibly all prostate cancers, can be subtyped to only three distinct groups. This is a major discovery in the field, which has allowed collaborations with other nationally prominent prostate cancer research teams, including at the University of Michigan and UCLA. Early versions of this work have also been presented at AUA, AACR, and PCF national conferences. These findings were reported in ***Cancer Research***, which can be accessed by PMC version of full text manuscript in order to facilitate share the results from this study in public domain. I also gave an oral presentation at the 2015 The Western Section American Urological Association (AUA) Meeting and the 2016 American Urological Association (AUA) Summer Research Conference.

What do you plan to do during the next reporting period to accomplish the goals?

Nothing to Report.

4. IMPACT

What was the impact on the development of the principal discipline(s) of the project?

I have made an important conceptual and clinically relevant advance by developing a novel method of characterizing prostate cancer using transcriptomic profiles. Consequently, this project is high impact and high reward, with potentially immediate opportunities to alter clinical practice if the classification scheme can be shown to have clinical utility. The new prostate cancer classification scheme I developed might improve prognostication of prostate cancer and enable the development of subtype-specific therapies and companion diagnostics. Using computational modeling, I have also identified a transcription factor, ONECUT2, which appears to be highly active in CRPC/Met tumors, but which has not been studied in PC, and therefore represents a first-in-field discovery. The comprehensive computational analyses and experimental interrogations of SAFB1 network in PC revealed the novel interactions of SAFB1, AR, and ONECUT2. In addition to this we could validate SAFB1 network can directly regulate UGT2B15 and UGT2B17 gene expression. Their expression seems to be coordinately regulated in many aggressive CRPC/Met patient samples. Collectively, these data suggest that SAFB1/AR/ONECUT2 network is a potential therapeutic target in CRPC.

What was the impact on other disciplines?
Nothing to Report.

What was the impact on technology transfer?
Nothing to Report.

What was the impact on society beyond science and technology?
Nothing to Report.

5. CHANGES/PROBLEMS

Changes in approach and reasons for change
Nothing to Report.

Actual or anticipated problems or delays and actions or plans to resolve them
Nothing to Report.

Changes that had a significant impact on expenditures
Nothing to Report.

Significant changes in use or care of human subjects, vertebrate animals, biohazards, and/or select agents
Nothing to Report.

6. PRODUCTS:

Publications, conference papers, and presentations

Journal publications.

1. **You S**, Knudsen BS, Erho N, Alshalalfa M, Takhar M, Ashab HA, Davicioni E, Karnes RJ, Klein EA, Den RB, Ross AE, Schaeffer EM, Garraway IP, Kim J, Freeman MR, Integrated classification of prostate cancer reveals a novel luminal subtype with poor outcome. Cancer Research, 2016; Jun 14. pii: canres.0902.2016. PMID: 27302169. Acknowledgement of federal support (Yes)

Books or other non-periodical, one-time publications.
Nothing to Report.

Other publications, conference papers, and presentations.

Poster presentation:

1. **You S**, Kim J, Freeman MR, An epigenomic pathway from cholesterol to intracrine androgen. The 2014 American Urological Association (AUA) Annual Meeting, held in Orlando, Florida, from May 16 to 21, 2014.
2. **You S**, Kim J, Freeman MR, Prostate Cancer Classification Using a Transcriptome Atlas. The Prostate Cancer Foundation (PCF) 21st Annual Scientific Retreat, held in Carlsbad, California, October from 23 to 25, 2014.
3. **You S**, Kim J, Freeman MR, Prostate cancer classification using a transcriptome atlas. American Association for Cancer Research (AACR) Special Conference. 2015.
4. **You S**, Erho N, Alshalalfa M, Takhar M, Ashab HA, Davicioni E, Karnes J, Klein EA, Den RB, Garraway IP, Knudsen BS, Kim J, Freeman MR, Three intrinsic subtypes of prostate cancer with distinct pathway activation profiles differ in prognosis and treatment response. The Prostate Cancer Foundation (PCF) 22nd Annual Scientific Retreat. 2015.
5. **You S**, Knudsen BS, Erho N, Alshalalfa M, Takhar M, Ashab HA, Davicioni E, Karnes J, Klein EA, Den RB, Garraway IP, Kim J, Freeman MR, Three intrinsic subtypes of prostate cancer with distinct pathway activation profiles differ in prognosis and treatment response. The 2016 American Urological Association (AUA) Annual Meeting, held in San Diego, California, May, 2016.

Lecture:

1. **You S**, Kim J, Introduction to Bioinformatics. The Urologic Oncology Program, held in Cedars-Sinai Medical Center, Los Angeles, California, March 10, 2015.

Oral Presentation:

1. **You S**, An epigenomic pathway from cholesterol to intracrine androgen. The 2015 Western Section American Urological Association (AUA) Meeting, held in Palm Springs, California, October 25, 2015.
2. **You S**, Integrated classification of prostate cancer reveals a novel luminal subtype with poor outcome. The 2016 American Urological Association (AUA) Summer Research Conference, held in Linthicum, Maryland, July 16, 2016.

Website(s) or other Internet site(s)

Nothing to Report.

Technologies or techniques

Nothing to Report.

Inventions, patent applications, and/or licenses

Patent applications:

1. **You S**, Freeman MR, Kim J, Knudsen B, Method of Diagnosing and Treating Prostate Cancer, Reference Number: 065472-000582PR00, 2015.
2. Rotinen M, **You S**, Murali R, Freeman MR, Agent for Treating Castration Resistant Prostate Cancer, Reference Number: 065472-000593PR00, 2015.

Other Products

Nothing to Report.

7. PARTICIPANTS & OTHER COLLABORATING ORGANIZATIONS

What individuals have worked on the project?

Name:	Sungyong You
Project Role:	Principal Investigator
Researcher Identifier:	yousung1
Nearest person month worked:	5

Contribution to Project:	Dr. You has performed all the works in computational analysis and experiments
Funding Support:	The Urology Care Foundation Research Scholar Program

Has there been a change in the active other support of the PD/PI(s) or senior/key personnel since the last reporting period?
Nothing to Report.

What other organizations were involved as partners?
Nothing to Report.

8. SPECIAL REPORTING REQUIREMENTS: Nothing to Report.

9. APPENDICES: A copy of journal publication

Cancer Research

Integrated Classification of Prostate Cancer Reveals a Novel Luminal Subtype with Poor Outcome

Sungyong You[1], Beatrice S. Knudsen[1], Nicholas Erho[2], Mohammed Alshalalfa[2], Mandeep Takhar[2], Hussam Al-deen Ashab[2], Elai Davicioni[2], R. Jeffrey Karnes[3], Eric A. Klein[4], Robert B. Den[5], Ashley E. Ross[6], Edward M. Schaeffer[6], Isla P. Garraway[7], Jayoung Kim[1], and Michael R. Freeman[1]

Abstract

Prostate cancer is a biologically heterogeneous disease with variable molecular alterations underlying cancer initiation and progression. Despite recent advances in understanding prostate cancer heterogeneity, better methods for classification of prostate cancer are still needed to improve prognostic accuracy and therapeutic outcomes. In this study, we computationally assembled a large virtual cohort ($n = 1{,}321$) of human prostate cancer transcriptome profiles from 38 distinct cohorts and, using pathway activation signatures of known relevance to prostate cancer, developed a novel classification system consisting of three distinct subtypes (named PCS1–3). We validated this subtyping scheme in 10 independent patient cohorts and 19 laboratory models of prostate cancer, including cell lines and genetically engineered mouse models. Analysis of subtype-specific gene expression patterns in independent datasets derived from luminal and basal cell models provides evidence that PCS1 and PCS2 tumors reflect luminal subtypes, while PCS3 represents a basal subtype. We show that PCS1 tumors progress more rapidly to metastatic disease in comparison with PCS2 or PCS3, including PSC1 tumors of low Gleason grade. To apply this finding clinically, we developed a 37-gene panel that accurately assigns individual tumors to one of the three PCS subtypes. This panel was also applied to circulating tumor cells (CTC) and provided evidence that PCS1 CTCs may reflect enzalutamide resistance. In summary, PCS subtyping may improve accuracy in predicting the likelihood of clinical progression and permit treatment stratification at early and late disease stages. *Cancer Res; 76(17); 1–11. ©2016 AACR.*

Introduction

Prostate cancer is a heterogeneous disease. Currently defined molecular subtypes are based on gene translocations (1, 2), gene expression (3, 4), mutations (5–8), and oncogenic signatures (9, 10). In other cancer types, such as breast cancer, molecular classifications predict survival and are routinely used to guide treatment decisions (11, 12). However, the heterogeneous nature of prostate cancer, and the relative paucity of redundant genomic alterations that drive progression, or that can be used to assess likely response to therapy, have hindered attempts to develop a classification system with clinical relevance (13).

Recently, molecular lesions in aggressive prostate cancer have been identified. For example, overexpression of the androgen receptor (*AR*) due to gene amplification has been observed in castration-resistant prostate cancer (CRPC) (14). Presence of AR variants (AR-V) that do not require ligand for activation have been reported in a large percentage of CRPCs and have been correlated with resistance to AR-targeted therapy (15). The oncogenic function of enhancer of zeste homolog 2 (*EZH2*) was found in cells of CRPC, and recurrent mutations in the speckle-type POZ protein (*SPOP*) gene occur in approximately 15% of prostate cancers (16, 17). Expression signatures related to these molecular lesions have also been developed to predict patient outcomes. While, in principle, signature-based approaches could be used independently in small cohorts (4, 10), there is a potential for an increase in diagnostic or prognostic accuracy if signatures reflecting gene expression perturbations relevant to prostate cancer could be applied to large cohorts containing thousands of clinical specimens.

Here we present the results of an integrated analysis of an unprecedentedly large set of transcriptome data, including from over 4,600 clinical prostate cancer specimens. This study revealed that RNA expression data can be used to categorize prostate cancer

[1]Division of Cancer Biology and Therapeutics, Departments of Surgery & Biomedical Sciences, Samuel Oschin Comprehensive Cancer Institute, Cedars-Sinai Medical Center, Los Angeles, California. [2]GenomeDx Biosciences Inc., Vancouver, British Columbia, Canada. [3]Department of Urology, Mayo Clinic, Rochester, Minnesota. [4]Glickman Urological and Kidney Institute, Cleveland Clinic, Cleveland, Ohio. [5]Department of Radiation Oncology, Jefferson Medical College of Thomas Jefferson University, Philadelphia, Pennsylvania. [6]Department of Urology, Northwestern University, Chicago, Illinois. [7]Department of Urology, David Geffen School of Medicine at UCLA, Los Angeles, California.

Note: Supplementary data for this article are available at Cancer Research Online (http://cancerres.aacrjournals.org/).

Corresponding Author: Michael R. Freeman, Division of Cancer Biology and Therapeutics, Departments of Surgery & Biomedical Sciences, Samuel Oschin Comprehensive Cancer Institute, Cedars-Sinai Medical Center, Davis 5072, 8700 Beverly Blvd, Los Angeles, CA 90048. Phone: 617-721-6427; Fax: 310-423 0139; E-mail: Michael.Freeman@cshs.org

doi: 10.1158/0008-5472.CAN-16-0902

tumors into 3 distinct subtypes, based on molecular pathway representation encompassing molecular lesions and cellular features related to prostate cancer biology. Application of this subtyping scheme to 10 independent cohorts and a wide range of preclinical prostate cancer models strongly suggest that the subtypes we define originate from inherent differences in prostate cancer origins and/or biological features. We provide evidence that this novel prostate cancer classification scheme can be useful for detection of aggressive tumors using tissue as well as blood from patients with progressing disease. It also provides a starting point for development of subtype-specific treatment strategies and companion diagnostics.

Materials and Methods

Merging transcriptome datasets and quality control

To assemble a merged dataset from diverse microarray and high-throughput sequencing platforms, we applied a median-centering method followed by quantile scaling (MCQ; ref. 18). Briefly, each dataset was normalized using the quantile method (19). Probes or transcripts were assigned to unique genes by mapping NCBI entrez gene IDs. Redundant replications for each probe and transcript were removed by selecting the one with the highest mean expression. Log2 intensities for each gene were centered by the median of all samples in the dataset. Each of the matrices was then transformed into a single vector. The vectors for the matrices were scaled by the quantile method to avoid a bias toward certain datasets or batches with large variations from the median values. These scaled vectors were transformed back into the matrices. Finally, the matrices were combined by matching the gene IDs in the individual matrices, resulting in a merged dataset of 2,115 samples by 18,390 human genes. To evaluate the MCQ-based normalization strategy, we applied the XPN (cross platform normalization; ref. 20) method to the same datasets and compared it with the merged data from MCQ. Multidimensional scaling (MDS) between samples was performed to assess batch effects. The same MCQ approach with the quantile method, or the single channel array normalization (SCAN) method (21), was also applied for normalization and batch correction of data from the independent cohorts.

Computing pathway activation score

We used the *Z*-score method to quantify pathway activation (22). Briefly, the *Z*-score was defined by the difference between the error-weighted mean of the expression values of the genes in a gene signature and the error-weighted mean of all genes in a sample after normalization. *Z*-scores were computed using each signature in the signature collection for each of the samples, resulting in a matrix of pathway activation scores.

Determination of the optimal number of clusters

Non-negative matrix factorization (NMF) clustering with a consensus approach is useful to elucidate biologically meaningful classes (23). Thus, we applied the consensus NMF clustering method (24) to identify the optimal number of clusters. NMF was computed 100 times for each rank k from 2 to 6, where k was a presumed number of subtypes in the dataset. For each k, 100 matrix factorizations were used to classify each sample 100 times. The consensus matrix with samples was used to assess how consistently sample-pairs cluster together. We then computed the cophenetic coefficients and silhouette scores for each k, to quantitatively assess global clustering robustness across the consensus matrix. The maximum peak of the cophenetic coefficient and silhouette score plots determined the optimal number of clusters.

Classification using a 14-pathway classifier

We constructed a classifier, where a set of predictors consists of 14 pathways, using a naïve Bayes machine learning algorithm. For training the classifier, we used the pathway activation scores and subtype labels of the result of the NMF clustering process. We then computed the misclassification rate using stratified 10-fold cross validation. To assess performance, we adopted a 3-class classification as a 2-class classification (e.g., PCS1 vs. others) and computed the average area under the receiver operating characteristic (ROC) curves from all 3 of 2-class classifications. Finally, we applied the 14-pathway classifier to assign subtypes to the specimens.

Identifying subtype-enriched genes

Wilcoxon rank-sum test and subsequent false discovery rate (FDR) correction with Storey's method (25) were employed to identify differentially expressed genes between the subtypes. Genes were selected with FDR < 0.001 and fold change ≥ 1.5, resulting in 428 subtype-enriched genes (SEG).

Development of a 37-gene diagnostic panel

A random forest machine learning algorithm was employed to develop a diagnostic gene panel. For parameter estimation and training the model, we used the merged dataset. Initially, the model comprised of the 428 SEGs as a set of predictors and subtype label of the merged dataset was used as a response variable for model training. To verify the optimal leaf size, we compared the mean squared errors (MSE) obtained by classification of leaf sizes of 1 to 50 with 100 trees, resulting in an optimal leaf size of 1 for model training. We then permuted the values for each gene across every sample and measured how much worse MSE became after the permutation. Imposing a cutoff of importance score at 0.5, we selected the 37 genes for subtyping. From the computation of MSE growing 100 trees on 37 genes and on the 428 SEGs, the 37 genes we chose gave the same MSE as the full set of 428 genes. ROC curve analyses and 10-fold cross-validation were also conducted to assess the performance of a classification ensemble.

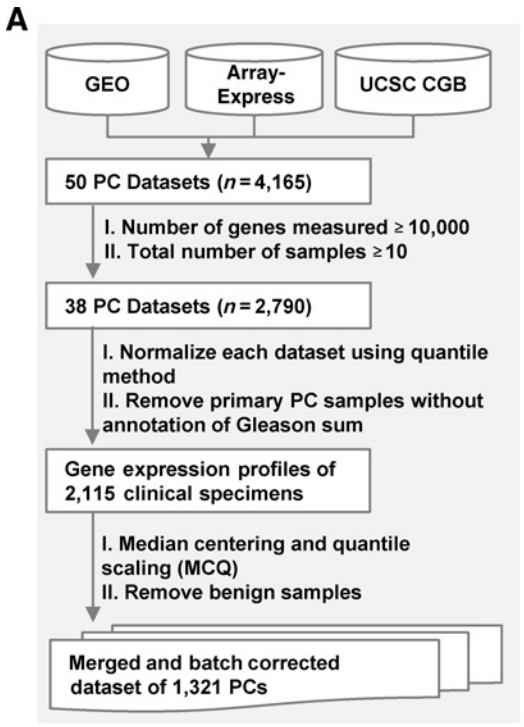
Statistical analysis

We performed principal component analysis (PCA) and MDS for visualizing the samples to assess their distribution using pathway activation profiles. Wilcoxon rank-sum statistics were used to test for significant differences in pathway activation scores between the subtypes. Kaplan–Meier analysis, Cox proportional hazard regression, and the χ^2 test were performed to examine the relationship(s) between clinical variables and subtype assignment. The OR test using dichotomized variables was conducted to investigate relationships between different subtyping schemes. The MATLAB package (Mathworks) and the R package (v.3.1 http://www.r-project.org/) were used for all statistical tests.

Results

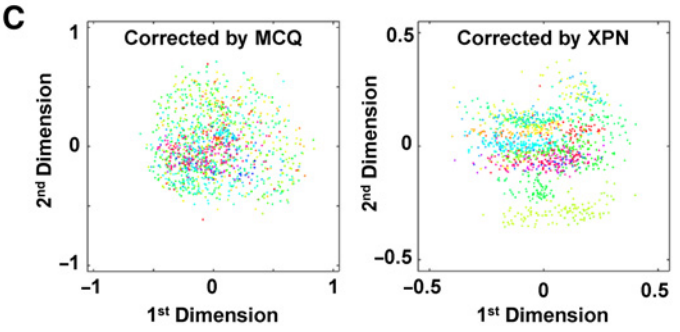
A prostate cancer gene expression atlas

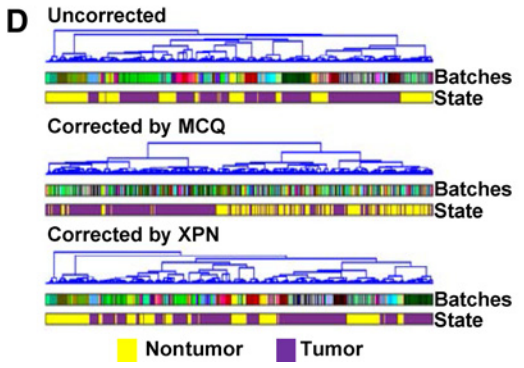
To achieve adequate power for a robust molecular classification of prostate cancer, we initially collected 50 prostate cancer



B

Disease status	Number of samples	Sample composition
Benign	794	Benign prostate tissue
GS < 7	328	Primary PC with Gleason sum < 7
GS = 7	530	Primary PC with Gleason sum = 7
GS > 7	203	Primary PC with Gleason sum > 7
CRPC/Met	260	metastatic or castration-resistant PC





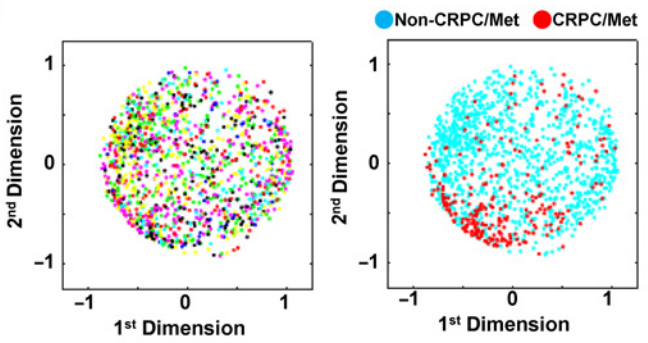


Figure 1.

Integration of prostate cancer transcriptome and quality control. **A,** schematic showing the process of collecting and merging prostate cancer transcriptomes. **B,** clinical composition of 2,115 prostate cancer cases. **C,** MDS of merged expression profiles after MCQ or XPN correction in the DISC cohort. Dots with different colors represent different batches or datasets. **D,** hierarchical clustering illustrates the sample distribution of uncorrected (top), corrected by MCQ (middle), and corrected by XPN (bottom). Different colors on "Batches" rows represent different batches or datasets from the individual studies. **E,** MDS of pathway activation profiles in the DISC cohort shows distribution of the samples from same batches. Dots with different colors represent different batches or datasets.

datasets from three public databases: Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo), ArrayExpress (http://www.ebi.ac.uk/arrayexpress), and the UCSC Cancer Genomics Browser (https://genome-cancer.ucsc.edu) and selected 38 datasets (Supplementary Table S1), in which the numbers of samples are larger than 10 and where over 10,000 genes were measured (Fig. 1A). This collection contains datasets consisting of 2,790 expression profiles of benign prostate tissue, primary tumors, and metastatic or CRPC (CRPC/Met; Fig. 1B). We then removed a subset of samples with ambiguous clinical information and generated a single merged dataset by cross study normalization, based on median-centering and the quantile normalization method (MCQ; ref. 18). The merged dataset consists of 1,321 tumor specimens that we named the Discovery (DISC) cohort. The merged gene expression profiles showed a significant reduction of systematic, dataset-specific bias in comparison with the same dataset corrected by the XPN method, which is also used for merging data from different platforms (20) (Fig. 1C). Biological differences between tumors and benign tissues were also maintained while minimizing batch effects (Fig. 1D).

As validation datasets, we assembled another collection of 12 independent cohorts consisting of 2,728 tumors from primary and CRPC/Met samples (Table 1). From this collection, 3 datasets, the Swedish watchful waiting cohort (SWD), the Emory cohort (EMORY), and the Health Study Prostate Tumor cohort (HSPT), were obtained from GEO. The gene expression

Table 1. List of independent cohorts for validation of the subtypes

Cohort name	Number of samples	Disease status	Available clinical outcomes	Data from GRID	Abbreviation	PubMed ID
Swedish Watchful-Wainting Cohort	281	Localized	OS	No	SWD	20233430
The Cancer Genome Anatomy	333	Localized	N.A.	No	TCGA	26000489
Emory University	106	Localized	N.A.	No	EMORY	24713434
Health Professionals Follow-up Study and Physicians' Health Study Prostate Tumor Cohort	264	Localized	N.A.	No	HSPT	25371445
Stand Up To Cancer/Prostate Cancer Foundation Dream Team Cohort	118	CRPC/Met	N.A.	No	SU2C	26000489
Mayo Clinic Cohort 1	545	Localized	PMS, TMP, PCSM	Yes	MAYO1	23826159
Mayo Clinic Cohort 2	235	Localized	PMS, TMP, PCSM	Yes	MAYO2	23770138
Thomas Jefferson University cohort	130	Localized	PMS, TMP, PCSM	Yes	TJU	25035207
Cleveland Clinic Foundation Cohort	182	Localized	PMS, TMP, PCSM	Yes	CCF	25466945
Memorial Sloan Kettering Cancer Center cohort	131	Localized	PMS, PCSM	Yes	MSKCC	20579941
Erasmus Medical Centre Cohort	48	Localized	PMS, PCSM	Yes	EMC	23319146
Johns Hopkins Medicine Cohort	355	Localized	PMS, TMP, PCSM	Yes	JHM	25466945

Abbreviations: N.A., not available; OS, overall survival; PMS, progression to metastatic state; PCSM, PC-specific mortality; TMP, time-to-metastatic progression.

profiles and clinical annotations of The Cancer Gnome Atlas (TCGA) cohort of 333 prostate cancer and SU2C/PCF Dream Team cohort (SU2C) of 118 CRPC/Mets were obtained from cBioPortal (http://www.cbioportal.org/). Seven additional cohorts were obtained from the Decipher GRID database (GRID). The expression datasets from the GRID were generated using a single platform, the Affymetrix Human Exon 1.0 ST Array, using primary tumors for the purpose of developing outcomes and treatment response signatures. We used these 7 cohorts to investigate associations of clinical outcomes with subtype assignment in this study.

Pathway activations describing prostate cancer biology

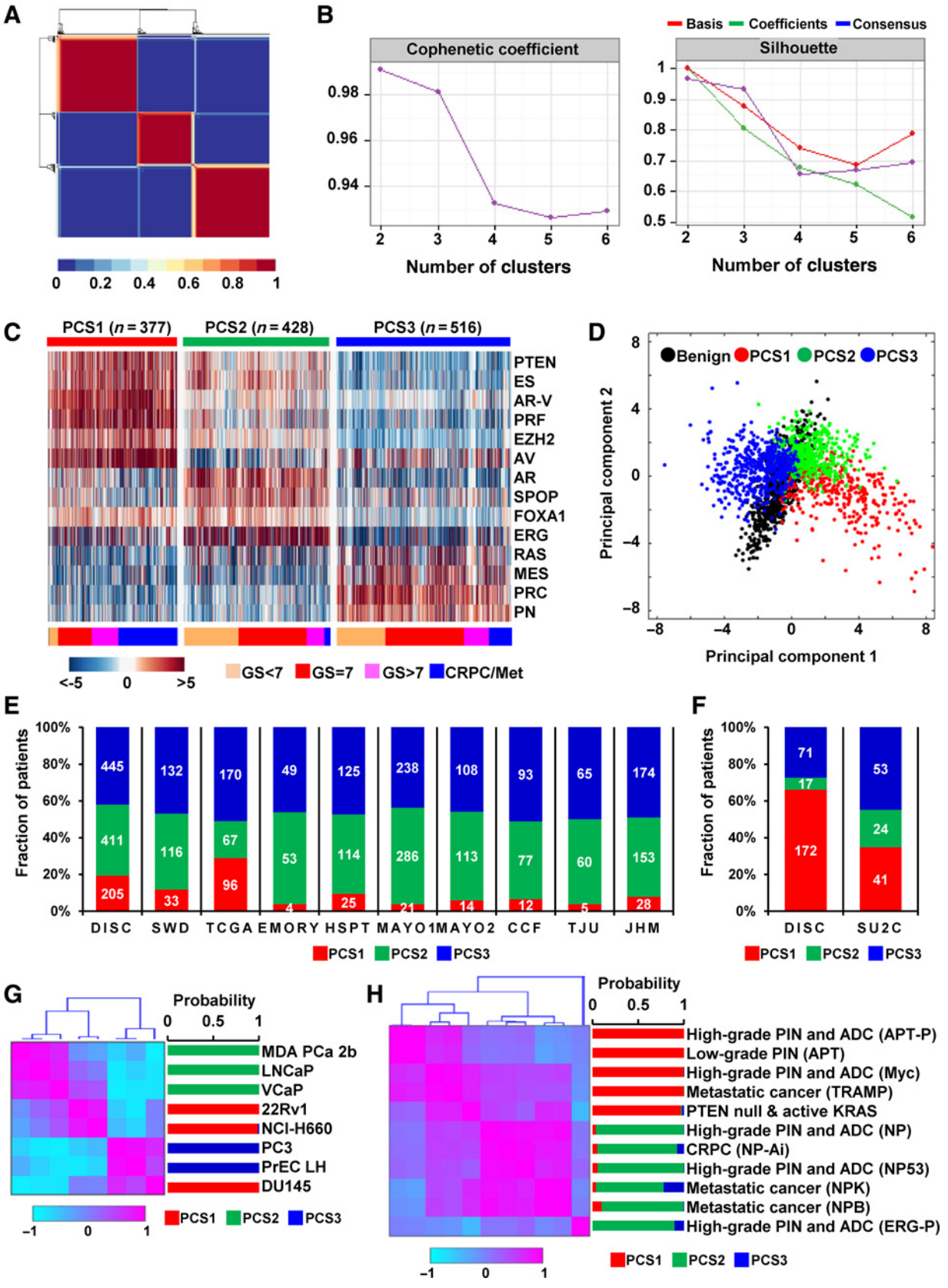
Recent studies have demonstrated the advantage of pathway-based analysis in clinical stratification for prostate and other cancer types (10, 26, 27), However, to date, there has been no study of prostate cancer using pathway activation profiles in which thousands of patient specimens were used. In addition, the utility of recently characterized molecular lesions such as *AR* amplification/overexpression, *AR-V* expression, transcriptional activation of *EZH2* and forkhead box A1 (*FOXA1*), and *SPOP* mutation have not been fully exploited for classification. Therefore, we employed 22 pathway activation gene expression signatures encompassing prostate cancer–relevant signaling and genomic alterations (Supplementary Tables S2 and S3) in the DISC cohort ($n = 1{,}321$). These were ultimately collapsed into 14 pathway signatures that were grouped into 3 categories: (i) prostate cancer–relevant signaling pathways, including activation of AR, AR-V, EZH2, FOXA1, and rat sarcoma viral oncogene homolog (RAS) and inactivation by polycomb repression complex 2 (PRC); (ii) genetic and genomic alterations, including mutation of *SPOP*, *TMPRSS2–ERG* fusion (ERG), and deletion of PTEN; and (iii) biological features related to aggressive prostate cancer progression, including stemness (ES), cell proliferation (PRF), epithelial–mesenchymal transition (MES), proneural (PN), and aggressive prostate cancer with neuroendocrine differentiation (AV). Pathway activation scores were computed in each specimen in the DISC cohort using the *Z*-score method (22). The conversion of gene expression to pathway activation showed a further reduction of batch effects, while preserving biological differences that are particularly evident in the clustering of metastatic and non-metastatic samples (Fig. 1E).

Identification and validation of molecular subgroups

We performed unsupervised clustering based on consensus NMF clustering (24) using the 14 pathway activation profiles in the DISC cohort. A consensus map of the NMF clustering results shows clear separation of the samples into three clusters (Fig. 2A). To identify the optimal number of clusters and to assess robustness of the clustering result, we computed the cophenetic coefficient and silhouette score using different numbers of clusters (2–6). These results indicate that 3 clusters is a statistically optimal representation of the data (Fig. 2B). A heatmap of 3 sample clusters demonstrates highly consistent pathway activation patterns within each group (Fig. 2C). These analyses suggest that the clusters correspond to three prostate cancer subtypes. We compared the magnitude of activation of each pathway across the 3 clusters evident in Fig. 2C using the Wilcoxon rank-sum test for pairwise comparisons (Supplementary Fig. S1). The PCS1 subtype exhibits high activation scores for EZH2, PTEN, PRF, ES, AV, and AR-V pathways. In contrast, ERG pathway activation predominates in PCS2, which is also characterized by high activation of AR, FOXA1, and SPOP. PCS3 exhibits high activation of RAS, PN, MES, while AR and AR-V activation are low.

High enrichment of PRC and low AR within PCS3 raises the question of whether this subtype is an artifact of contaminating nontumor tissues. However, PCA demonstrates that samples in PCS3 are as distinct from benign tissues as samples in the other subtypes (Fig. 2D). To further confirm the difference from benign tissue, we made use of a gene signature shown to discriminate benign prostate tissue from cancer in a previous study (28) and found a significant difference ($P < 0.001$) in all the tumors in the subtypes compared with benign tissues (Supplementary Fig. S2). These results demonstrate that prostate cancers retain distinct gene expression profiles between subtypes, which are not related to the amount of normal tissue contamination.

To validate the PCS classification scheme, a 14-pathway classifier was developed using a naïve Bayes machine learning algorithm (see details in Materials and Methods). This classifier was applied to 9 independent cohorts of localized tumors (i.e., SWD, TCGA, EMORY, HSPT, MAYO1/2, CCF, TJU, and JHM) and the SU2C cohort of CRPC/Met tumors. Out of these 10 independent cohorts, 5 cohorts (i.e., MAYO1/2, TJU, CCF, and JHM) were from the GRID (Fig. 2E; Table 1; ref. 29). The 14-pathway classifier reliably categorized tumors in the DISC cohort into 3 subtypes, with an average classification performance = 0.89 ($P < 0.001$). The


A
0 0.2 0.4 0.6 0.8 1
B
Cophenetic coefficient
0.98
0.96
0.94
2 3 4 5 6
Number of clusters
Basis Coefficients Consensus
Silhouette
1
0.9
0.8
0.7
0.6
0.5
2 3 4 5 6
Number of clusters
C
PCS1 (n = 377)
PCS2 (n = 428)
PCS3 (n = 516)
PTEN
ES
AR-V
PRF
EZH2
AV
AR
SPOP
FOXA1
ERG
RAS
MES
PRC
PN
<-5 0 >5
GS<7 GS=7 GS>7 CRPC/Met
D
Benign PCS1 PCS2 PCS3
Principal component 2
8 4 0 −4 −8
−8 −4 0 4 8
Principal component 1
E
Fraction of patients
100% 80% 60% 40% 20% 0%
445 411 205
132 116 33
170 67 96
49 53 4
125 114 25
238 286 21
108 113 14
93 77 12
65 60 5
174 153 28
DISC SWD TCGA EMORY HSPT MAYO1 MAYO2 CCF TJU JHM
PCS1 PCS2 PCS3
F
Fraction of patients
100% 80% 60% 40% 20% 0%
71 17 172
53 24 41
DISC SU2C
PCS1 PCS2 PCS3
G
Probability
0 0.5 1
MDA PCa 2b
LNCaP
VCaP
22Rv1
NCI-H660
PC3
PrEC LH
DU145
−1 0 1
PCS1 PCS2 PCS3
H
Probability
0 0.5 1
High-grade PIN and ADC (APT-P)
Low-grade PIN (APT)
High-grade PIN and ADC (Myc)
Metastatic cancer (TRAMP)
PTEN null & active KRAS
High-grade PIN and ADC (NP)
CRPC (NP-Ai)
High-grade PIN and ADC (NP53)
Metastatic cancer (NPK)
Metastatic cancer (NPB)
High-grade PIN and ADC (ERG-P)
−1 0 1
PCS1 PCS2 PCS3

3 subtypes were identified in all cohorts. Their proportions were similar across the localized disease cohorts, demonstrating the consistency of the classification algorithm across multiple practice settings (Fig. 2E). The 2 cohorts consisting of CRPC/Met tumors (DISC and SU2C) showed some differences in the frequency of PCS1 and PCS3; the most frequent subtype in the DISC CRPC/Met cohort was PCS1 (66%), while the most frequent subtype in SU2C was PCS3 (45%; Fig. 2F). PCS2 was the minor subtype in both CRPC/Met cohorts.

To determine whether the PCS classification is relevant to laboratory models of prostate cancer, we analyzed 8 human prostate cancer cell lines from The Cancer Cell Line Encyclopedia (CCLE; GSE36133; ref. 30) and 11 prostate cancer mouse models (31, 32). There are two datasets for mouse models. The first dataset (GSE53202) contains transcriptome profiles of 13 genetically engineered mouse models, including normal epithelium (i.e., wild-type), low-grade PIN (i.e., Nkx3.1 and APT), high-grade PIN, and adenocarcinoma (i.e., APT-P, APC, Myc, NP, Erg-P, and NP53), CRPC (i.e., NP-Ai), and metastatic prostate cancer (i.e., NPB, NPK, and TRAMP). Because of no available data for samples without drug treatment, the Nkx3.1 and APC models were excluded from this analysis. The second dataset (GSE34839) contains transcriptome profiles from mice with *PTEN*-null/*KRAS* activation mutation-driven high-grade, invasive prostate cancer and mice with only the *PTEN*-null background. This analysis revealed that all 3 prostate cancer subtypes were represented in the 8 human prostate cancer cell lines (Fig. 2G), while only 2 subtypes (PCS1 and PCS2) were represented in the mouse models (Fig. 2H). This result provides evidence that the subtypes are recapitulated in genetically engineered mouse models and persist in human cancer cells in cell culture.

Evaluation of PCS subtypes in comparison with other subtypes

Several categorization schemes of prostate cancer have been described, based mostly on tumor-specific genomic alterations and in some cases with integration of transcriptomic and other profiling data (10, 29, 33). This prompted us to compare the PCS classification scheme with the genomic subtypes derived by TCGA (34), because comprehensive genomic categorization was recently made available (35). We also compared the PCS classification with the subtypes recently defined by Tomlins and colleagues from RNA expression data (29). The Tomlins subtyping scheme is defined using the 7 GRID cohorts (i.e., MAYO1/2, TJU, CCF, MSKCC, EMC, and JHM) that we used for validating the PCS system. The large number of cases in the 7 GRID cohorts (n = 1,626) is comparable with our DISC cohort in terms of heterogeneity and complexity. TCGA identified several genomic subtypes, named ERG, ETV1, ETV4, FLI1, SPOP, FOXA1, IDH1, and "other." Tomlins and colleagues described 4 subtypes based on microarray gene expression patterns that are related to several genomic aberrations [i.e., ERG^+, ETS^+, $SPINK1^+$, and triple negative ($ERG^-/ETS^-/SPINK1^-$)].

A comparison of the PCS categories with the TCGA genomic subtypes showed that the tumors classified as ERG, ETV1/4, SPOP, FOXA1, and "other" were present across all the PCS categories in the TCGA dataset ($n = 333$; Fig. 3A). SPOP cancers were enriched in PCS1 (OR: 3.53), while PCS2 tumors were overrepresented in TCGA/ERG cancers (OR: 1.82) and TCGA/"other" cancers were enriched in PCS3 (OR: 1.79; Fig. 3B). In the GRID cohorts, we observed all PCS categories in all classification groups as defined by Tomlins and colleagues (Fig. 3C and D). We found a high frequency of the Tomlins/ERG^+ subtype in PCS2, but not in PCS1. PCS1 was enriched for Tomlins/ETS^+ and Tomlins/$SPINK1^+$ subtypes, while PCS3 was enriched for the triple-negative subtype but not the ERG^+ or ETS^+ subgroups. Finally, we compared the Tomlins classification method with the PCS classification using 5 of 7 GRID cohorts. PCS1 demonstrated significantly shorter metastasis-free survival compared with PCS2 and PCS3 ($P < 0.001$; Fig. 3E). In contrast, no difference in metastatic progression was seen among the Tomlins categories (Fig. 3F).

PCS1 contained the largest number of prostate cancers with GS $\geq$ 8 (Fig. 2C). Given the overall poorer outcomes seen in PCS1 tumors, we tested whether this result was simply a reflection of the enrichment of high-grade disease in this group (i.e., GS $\geq$ 8). For this analysis, we merged 5 GRID cohorts (i.e., MAYO1/2, TJU, CCF, and JHM) into a single dataset and separately analyzed low and high-grade disease. We observed a similarly significant ($P < 0.001$) association between subtypes and metastasis-free survival in GS $\leq$ 7 and in GS $\geq$ 8 (Fig. 3G). Thus, tumors in the PCS1 group exhibit the poorest prognosis, including in tumors with low Gleason sum score. Finally, in the DISC cohort, although CRPC/Met tumors were present in all PCS categories, PCS1 predominated (66%), followed by PCS3 (27%) and PCS2 (7%) tumors. To confirm whether this clinical correlation is replicated in individual cohorts, we also assessed association with time to metastatic progression, prostate cancer–specific mortality (PCSM), and overall survival (OS) in 5 individual cohorts in the GRID (i.e., MAYO1/2, CCF, TJU, and JHM) and in the SWD cohorts. PCS1 was seen to be the most aggressive subtype, consistent with the above results (Supplementary Fig. S3).

PCS categories possess characteristics of basal and luminal prostate epithelial cells

Prostate cancer may arise from oncogenic transformation of different cell types in glandular prostate epithelium (36–38). Breast cancers can be categorized into luminal and basal subtypes, which are associated with different patient outcomes (39). It is unknown whether this concept applies to human prostate cancer. To examine whether the 3 PCS categories are a reflection of different cell types, we identified 428 SEGs (SEG1–3; 86 for PCS1, 123 for PCS2, and 219 for PCS3; Supplementary Table S4) in each subtype. As expected, these genes are involved in pathways that are enriched in each subtype (Fig. 4A) and that define the perturbed cellular processes of the subtype. We then identified the cellular processes that are associated with the SEGs. Proliferation and lipid/steroid metabolism are characteristic of SEG1 and SEG2, while extracellular matrix organization, inflammation, and cell migration are characteristic of SEG3 (Fig. 4B). This result suggests

Figure 2.

Identification and validation of novel prostate cancer subtypes. **A,** consensus matrix depicts robust separation of tumors into three subtypes. **B,** changes of cophenetic coefficient and silhouette score at rank 2 to 6. **C,** pathway activation profiles of 1,321 tumors defines three prostate cancer subtypes. **D,** score plot of PCA for benign and three subtypes. **E** and **F,** the three subtypes were recognized in 10 independent cohorts. **G** and **H,** correlation of pathway activation profiles in 8 prostate cancer cell lines from the CCLE and 11 prostate cancer mouse models and probability from the pathway classifier.

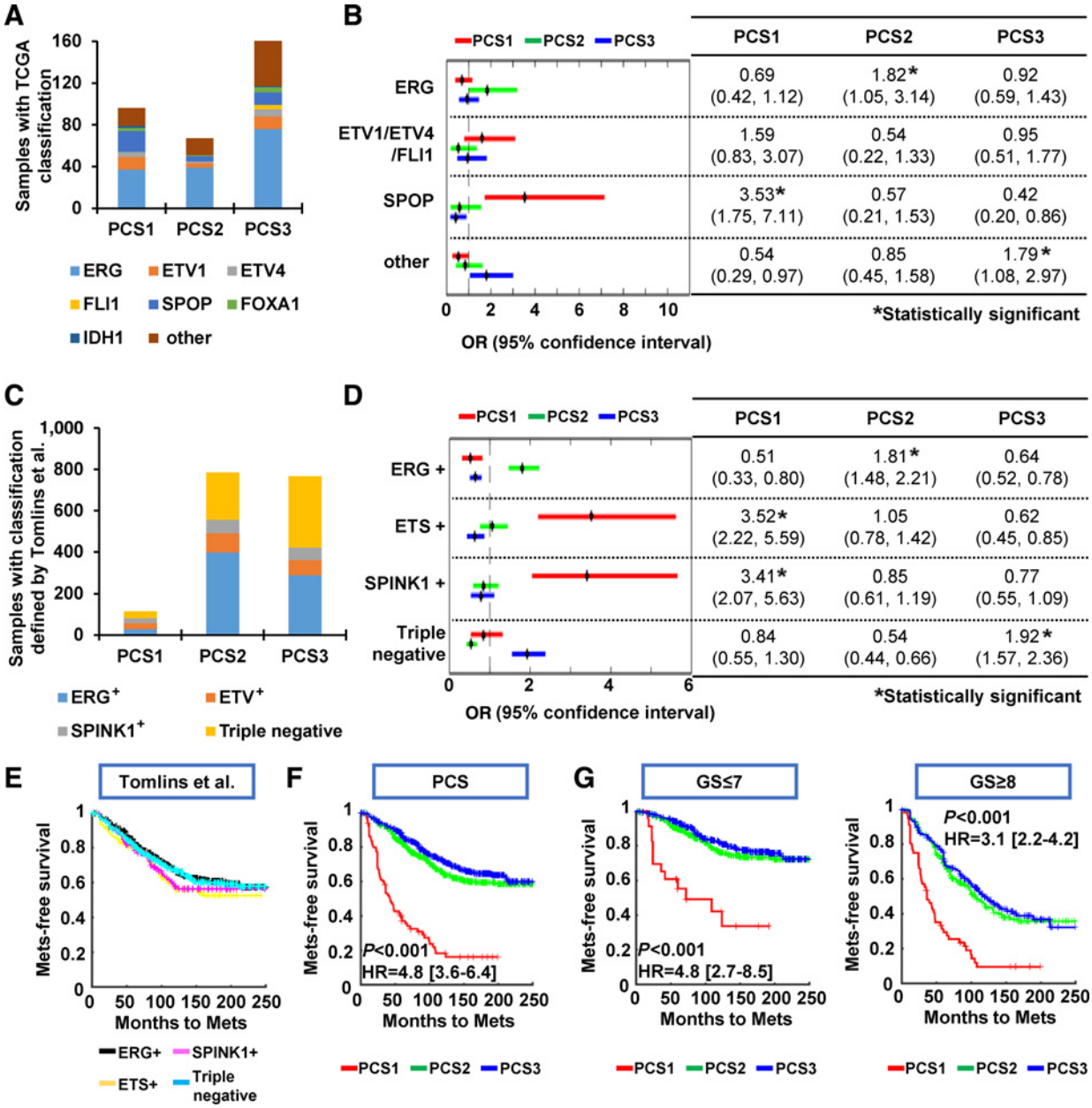


Figure 3.

Comparison of the PCS subtypes with previously described subtypes. **A,** distribution of TCGA tumors ($n = 333$) using the PCS subtypes compared with TCGA subtypes. **B,** relationship between PCS subtyping and TCGA subtypes. **C,** distribution of GRID tumors ($n = 1{,}626$) using PCS categories compared with Tomlins subtypes. **D,** relationship between PCS subtyping and Tomlins subtypes. **E** and **F,** association of metastasis-free survival using Tomlins subtypes and using the PCS subtypes in the GRID tumors. **G,** metastasis-free survival in tumors of GS $\leq$ 7 (left) and GS $\geq$ 8 (right).

that distinct biological functions are associated with the PCS categories.

To determine whether the PCS categories reflect luminal or basal cell types of the prostatic epithelium, we analyzed the mean expression of genes known to be characteristic of luminal (*EZH2*, *AR*, *MKI67*, *NKX3-1*, *KLK2/3*, and *ERG*) or basal (*ACTA2*, *GSTP1*, *IL6*, *KRT5*, and *TP63*) prostatic cells (Fig. 4C). We observed a strong association (FDR < 0.001; fold change > 1.5) between luminal genes and PCS1 and PCS2, and basal genes and PCS3. To verify this observation, we used two independent datasets derived from luminal and basal cells from human (40) and mouse (GSE39509; ref. 37) prostates. The assignment of a basal designation to PCS3 is further supported by the highly significant enrichment in PCS3, in comparison with the other two subtypes, of a recently described prostate basal cell signature derived from CD49f-Hi versus CD49f-Lo benign and malignant prostate epithelial cells (Fig. 4D; ref. 41). In addition, using the 14-pathway classifier, mouse basal tumors and human basal

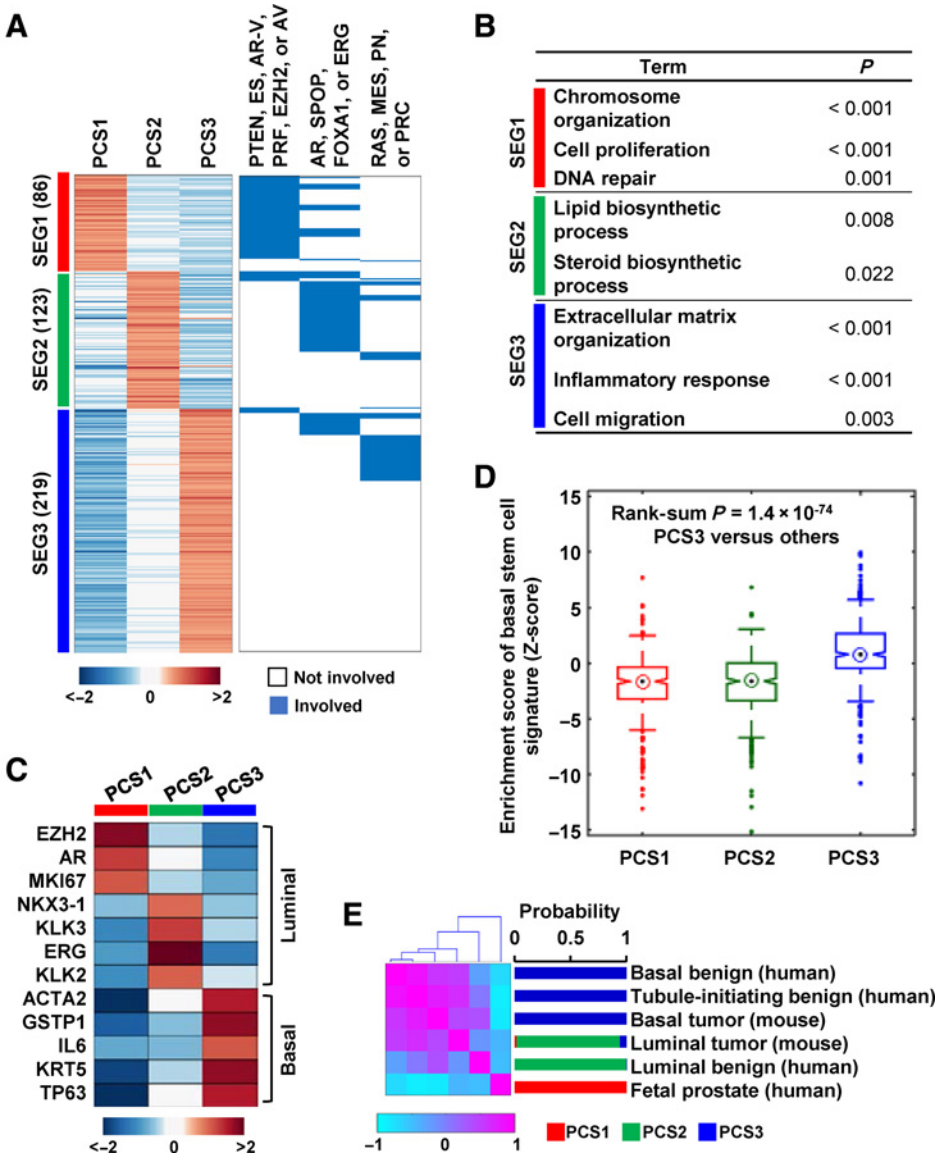


Figure 4.
Genes enriched in each of the three subtypes are associated with luminal and basal cell features. **A,** relative gene expression (left) and pathway inclusion (right) of SEGs are displayed. **B,** cellular processes enriched by each of the three SEGs ($P < 0.05$). **C,** expression of the luminal and basal markers in the three subtypes. **D,** enrichment of basal stem cell signature. **E,** correlation of pathway activities between samples from human and mouse prostate (left) and probability from the pathway classifier (right).

cells from benign tissues were classified as PCS3, while mouse luminal tumors and benign prostate human luminal cells were classified into PCS2 (Fig. 4E). These results are consistent with the conclusion that the PCS categories can be divided into luminal and basal subtypes.

A gene expression classifier for assignment to subtypes

Given the potential advantages of the PCS system to classify tumor specimens, we constructed a classifier that can be applied to an individual patient specimen in a clinical setting (Supplementary Fig. S4A). First, of 428 SEGs, 93 genes were selected on the basis of highly consistent expression patterns in 10 cohorts (i.e., SWD, TCGA, EMORY, HSPT, SU2C, MAYO1/2, CCF, TJU, and JHM). Second, using a random forest machine learning algorithm, we selected 37 genes with feature importance scores >0.5, showing a comparable level of error with the full model based on 428 SEGs (Supplementary Fig. S4B). Performance of the classifier was assessed in the GRID cohort (AUC = 0.97). The 37-gene panel displays significantly different expression patterns between the three subtypes in the DISC cohort (Fig. 5A).

The robust performance of the gene panel led us to determine whether it could be used to profile circulating tumor cells (CTC) from patients with CRPC. We analyzed single-cell RNA-seq data from 77 intact CTCs isolated from 13 patients (42). Prior to the clustering analysis to investigate the expression patterns of these CTC data, the normalized read counts as read-per-million (RPM) mapped reads were transformed on a log2 scale for each gene. The 77 CTCs were largely clustered into two groups using median-centered expression profiles corresponding to the 37-gene PCS panel by the hierarchical method (Fig. 5B). One group (GROUP I), consisting of 67 CTCs displays low expression of PCS1-enriched genes, while the other group (GROUP II) consisting of 10 CTCs has high expression of PCS1-enriched genes. In addition, we observed that PCS3-enriched genes in the panel were not detected or have very low expression changes across all CTCs as shown in the heatmap of Fig. 5B. The results suggest that CTCs can be

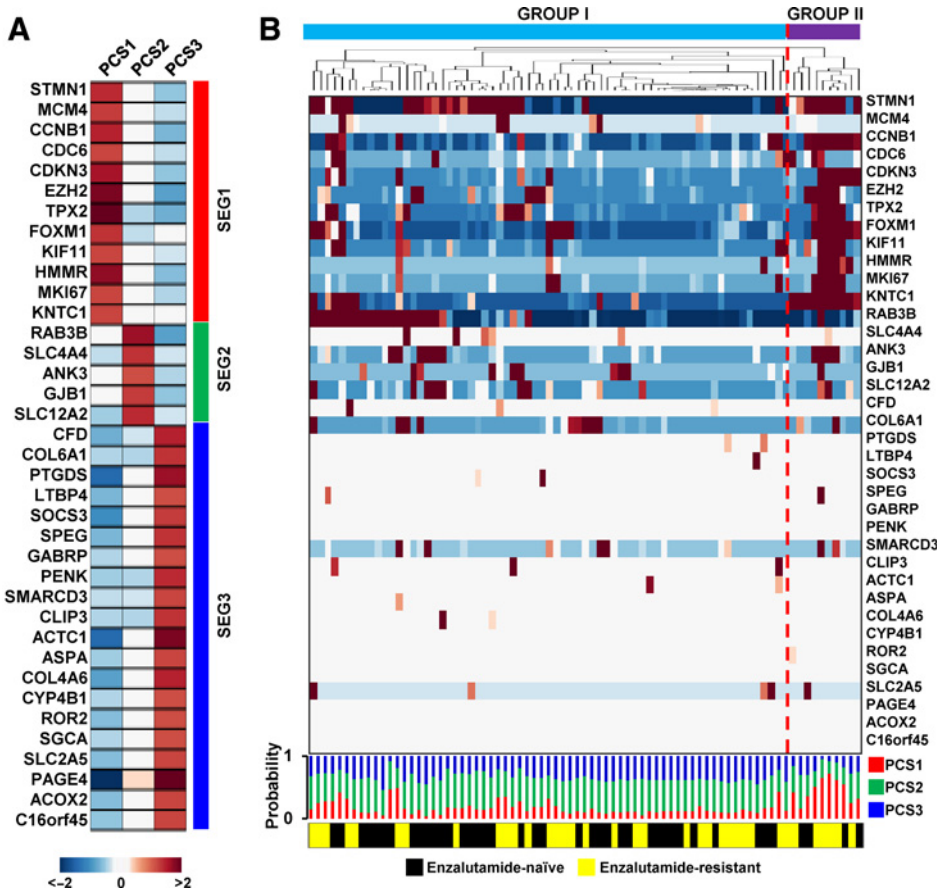


Figure 5.
A 37-gene classifier employed in patient tissues and CTCs. **A,** heatmap displays the mean expression pattern of the 37-gene panel in the three subtypes from the DISC cohort. **B,** hierarchical clustering of 77 CTCs obtained from CRPC patients by gene expression of the 37-gene panel. Bar plot in the bottom displays probability of PCS assignment from application of the classifier.

divided into two groups with the 37-gene PCS panel. Given this result, we hypothesized that the 37-gene classifier might assign CTCs to PCS1 or PCS2, consistent with the clustering result. The bar graph below the heatmap illustrates the probability of likelihood of PCS assignment, with the result that all the CTCs were assigned to PCS1 ($n = 12$) or PCS2 ($n = 65$), while no PCS3 CTCs were assigned on the basis of the largest probability score. By comparing with the CTC group assignment, 7 (70%) of 10 CTCs in the GROUP II were assigned to PCS1 by the 37-gene classifier and 62 (95%) of 65 CTCs in the GROUP I were assigned to PCS2 by the classifier. We then tested whether GROUP I and II exhibit any difference in terms of therapeutic responses. Of note, 5 of the 7 CTCs in GROUP II (OR: 1.74; 95% confidence interval: 0.49–6.06) were from patients whose cancer exhibited radiographic and/or PSA progression during enzalutamide therapy, suggesting that the 37-gene PCS panel can potentially identify patients with resistance to enzalutamide therapy.

Collectively, the results demonstrate that the 37-gene classifier has a potential to assign individual prostate cancers to PCS1 using both prostate tissues and blood CTCs, suggesting that the classifier can be applied to subtype individual prostate cancers using clinically relevant technology platforms (43, 44), including by noninvasive methods.

Discussion

In this study, we describe a novel classification system for prostate cancer, based on an analysis of over 4,600 prostate cancer specimens, which consists of only 3 distinct subtypes, designated PCS1, PCS2, and PCS3. PCS1 exhibits the highest risk of progression to advanced disease, even for low Gleason grade tumors. Although sampling methods across the cohorts we studied were different, classification into the 3 subtypes was reproducible. For example, the SWD cohort consists of specimens that were obtained by transurethral resection of the prostate rather than radical prostatectomy; however, subtype assignment and prognostic differences between the subtypes were similar to the other cohorts we examined (Supplementary Fig. S3J). Genes that are significantly enriched in the PCS1 category were highly expressed in the subset of CTCs (58%, 7 CTCs out of 12) from patients with enzalutamide-resistant tumors. This proportion of resistant cases in PCS1 CTCs is very high compared with PCS2 CTCs (8%, 5 CTCs out of 65). The characteristics of the PCS categories are summarized in Table 2.

Previously published prostate cancer classifications have defined subtypes largely based on the presence or absence of genomic alterations (e.g., *TMPRSS2-ERG* translocations). Tumors with ERG rearrangement (ERG^+) are overrepresented in PCS2; however, it is not the presence or absence of an ERG rearrangement that defines the PCS2 subtype, but rather ERG pathway activation features based on coordinate expression levels of genes in the pathway. Our findings provide evidence for biologically distinct forms of prostate cancer that are independent of Gleason grade, currently the gold standard for clinical decision-making. In addition, by comparing prognostic profiles between the PCS categories and the Tomlins and colleagues categories, prognostic

Table 2. Summary of PCS characteristics

Sample type	Features	PCS1	PCS2	PCS3
Patient tumors	Proportion	6%	47%	47%
	Pathology	Enriched GS $\geq$ 8	Enriched GS $\leq$ 7	Enriched GS $\leq$ 7
	Prognosis	Poor	Variable	Variable
	Subtypes- TCGA	SPOP	ERG	'other'
	Subtypes- Tomlins	ETS+, SPINK+	ERG+	Triple Negative
	Pathway signatures	AR-V, ES, PTEN, PRF, EZH2, NE	AR, FOXA1, SPOP, ERG	PRC, RAS, PN, MES
	Cell lineage	Luminal-like	Luminal-like	Basal-like
Patient CTCs	Proportion	16%	84%	0%
	Enzalutamide resistance	Yes (58%)	No (8%)	Unknown

information was evident only from the PCS classification scheme in the same cohort. Taken together, this indicates that the PCS classification is unique.

Although the current report has provided evidence that PCS classification can assign subtypes within groups of "indolent" as well as aggressive tumors, and in a wide range of preclinical models, it remains to be determined whether the PCS categories might be stable during tumor evolution in an individual patient. An interesting alternative possibility is that disease progression results in phenotypic diversification with respect to the PCS assignment. We have shown that preclinical model systems, including genetically engineered mouse models (GEMM), can be assigned with high statistical confidence to the PCS categories. We believe the simplest explanation for this finding is that these subtypes reflect distinct epigenetic features of chromatin that are potentially stable, even in the setting of genomic instability associated with advanced disease. This possibility needs to be formally tested. The human prostate cancer cell lines we evaluated could be assigned to all 3 subtypes; however, the GEMMs we tested could only be assigned to PCS1 and PCS2. This finding suggests that approximately 1 of 3 of human prostate cancers are not being modeled in widely used GEMMs. It should be feasible to generate mouse models for PCS3 through targeted genetic manipulation of pathways that are deregulated in PCS3 and through changing chromatin structure, such as by altering the activity of the PRC2 complex.

A major clinical challenge remains the early recognition of aggressive disease, in particular, due to the multifocal nature of prostate cancer (45). The classification scheme we describe predicts the risk of progression to lethal prostate cancer in patients with a diagnosis of low-grade localized disease (Fig. 3G). It is possible that in these cancers, pathway activation profiles are independent of Gleason grade and that pathways indicating high risk of progression are manifested early in the disease process and throughout multiple cancer clones in the prostate. In addition to predicting the risk of disease progression, PCS subtyping might also assist with the selection of drug treatment in advanced cancer by profiling CTCs in patient blood. With the 37-gene classifier we present here, it will be possible to assign individual tumors to PCS categories in a clinical setting. This new classification method may provide novel opportunities for therapy and clinical management of prostate cancer.

Disclosure of Potential Conflicts of Interest

N. Erho is a bioinformatics group lead at GenomeDx Biosciences Inc. M. Alshalalfa is a bioinformatician at GenomeDx Biosciences Inc. H. Al-deen Ashab is a data scientist at Genomedx Biosciences Inc. E. Davicioni has ownership interest (including patents) in GenomeDx Biosciences Inc. R.J. Karnes reports receiving other commercial research support from GenomeDx Biosciences Inc. E.A. Klein has received speakers bureau honoraria from GenomeDx Biosciences Inc. A.E. Ross has ownership interest (including patents) in GenomeDx Biosciences Inc. Mandeep Takhar is a bioinformatician at GenomeDX. No potential conflicts of interest were disclosed by the other authors.

Authors' Contributions

Conception and design: S. You, B.S. Knudsen, J. Kim, M.R. Freeman
Development of methodology: S. You, M. Alshalalfa, E.M. Schaeffer, J. Kim
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): S. You, B.S. Knudsen, E. Davicioni, R.J. Karnes, E.A. Klein, R.B. Den, A.E. Ross, E.M. Schaeffer, I.P. Garraway
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): S. You, N. Erho, M. Alshalalfa, H. Al-deen Ashab, E. Davicioni, E.M. Schaeffer, J. Kim, M.R. Freeman
Writing, review, and/or revision of the manuscript: S. You, B.S. Knudsen, N. Erho, M. Takhar, E. Davicioni, R.J. Karnes, E.A. Klein, R.B. Den, A.E. Ross, E.M. Schaeffer, I.P. Garraway, J. Kim, M.R. Freeman
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): S. You, M. Takhar, E. Davicioni, E.A. Klein, M.R. Freeman
Study supervision: S. You, M.R. Freeman

Acknowledgments

The authors are grateful to Drs. Francesca Demichelis, Felix Feng, and Benjamin Berman for helpful discussions during the course of this study.

Grant Support

This study was supported by grants from NIH (R01DK087806, R01CA143777, G20 RR030860, and 2P01 CA098912), the Department of Defense PCRP (W81XWH-14-1-0152; W81XWH-14-1-0273), the Urology Care Foundation Research Scholars Program, the Prostate Cancer Foundation Creativity Award, the Jean Perkins Foundation, and the Spielberg Family Discovery Fund.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received April 1, 2016; revised May 10, 2016; accepted May 25, 2016; published OnlineFirst June 14, 2016.

References

1. Perner S, Demichelis F, Beroukhim R, Schmidt FH, Mosquera JM, Setlur S, et al. TMPRSS2:ERG fusion-associated deletions provide insight into the heterogeneity of prostate cancer. Cancer Res 2006;66:8337–41.
2. Tomlins SA, Laxman B, Dhanasekaran SM, Helgeson BE, Cao X, Morris DS, et al. Distinct classes of chromosomal rearrangements create oncogenic ETS gene fusions in prostate cancer. Nature 2007;448:595–9.
3. Singh D, Febbo PG, Ross K, Jackson DG, Manola J, Ladd C, et al. Gene expression correlates of clinical prostate cancer behavior. Cancer Cell 2002;1:203–9.
4. Lapointe J, Li C, Higgins JP, van de Rijn M, Bair E, Montgomery K, et al. Gene expression profiling identifies clinically relevant subtypes of prostate cancer. Proc Natl Acad Sci U S A 2004;101:811–6.

5. Taylor BS, Schultz N, Hieronymus H, Gopalan A, Xiao Y, Carver BS, et al. Integrative genomic profiling of human prostate cancer. Cancer Cell 2010;18:11–22.
6. Grasso CS, Wu YM, Robinson DR, Cao X, Dhanasekaran SM, Khan AP, et al. The mutational landscape of lethal castration-resistant prostate cancer. Nature 2012;487:239–43.
7. Baca SC, Prandi D, Lawrence MS, Mosquera JM, Romanel A, Drier Y, et al. Punctuated evolution of prostate cancer genomes. Cell 2013;153:666–77.
8. Barbieri CE, Baca SC, Lawrence MS, Demichelis F, Blattner M, Theurillat JP, et al. Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer. Nat Genet 2012;44:685–9.
9. Tomlins SA, Mehra R, Rhodes DR, Cao X, Wang L, Dhanasekaran SM, et al. Integrative molecular concept modeling of prostate cancer progression. Nat Genet 2007;39:41–51.
10. Markert EK, Mizuno H, Vazquez A, Levine AJ. Molecular classification of prostate cancer using curated expression signatures. Proc Natl Acad Sci U S A 2011;108:21276–81.
11. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, et al. Molecular portraits of human breast tumours. Nature 2000;406:747–52.
12. Cancer Genome Atlas Research N. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. Nature 2008;455:1061–8.
13. Choudhury AD, Eeles R, Freedland SJ, Isaacs WB, Pomerantz MM, Schalken JA, et al. The role of genetic markers in the management of prostate cancer. Eur Urol 2012;62:577–87.
14. Sharma NL, Massie CE, Ramos-Montoya A, Zecchini V, Scott HE, Lamb AD, et al. The androgen receptor induces a distinct transcriptional program in castration-resistant prostate cancer in man. Cancer Cell 2013;23:35–47.
15. Watson PA, Arora VK, Sawyers CL. Emerging mechanisms of resistance to androgen receptor inhibitors in prostate cancer. Nat Rev Cancer 2015;15:701–11.
16. Xu K, Wu ZJ, Groner AC, He HH, Cai C, Lis RT, et al. EZH2 oncogenic activity in castration-resistant prostate cancer cells is Polycomb-independent. Science 2012;338:1465–9.
17. Blattner M, Lee DJ, O'Reilly C, Park K, MacDonald TY, Khani F, et al. SPOP mutations in prostate cancer across demographically diverse patient cohorts. Neoplasia 2014;16:14–20.
18. You S, Cho CS, Lee I, Hood L, Hwang D, Kim WU. A systems approach to rheumatoid arthritis. PLoS One 2012;7:e51508.
19. Bolstad BM, Irizarry RA, Astrand M, Speed TP. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics 2003;19:185–93.
20. Shabalin AA, Tjelmeland H, Fan C, Perou CM, Nobel AB. Merging two gene-expression studies via cross-platform normalization. Bioinformatics 2008;24:1154–60.
21. Piccolo SR, Sun Y, Campbell JD, Lenburg ME, Bild AH, Johnson WE. A single-sample microarray normalization method to facilitate personalized-medicine workflows. Genomics 2012;100:337–44.
22. Levine DM, Haynor DR, Castle JC, Stepaniants SB, Pellegrini M, Mao M, et al. Pathway and gene-set activation measurement from mRNA expression data: the tissue distribution of human pathways. Genome Biol 2006;7:R93.
23. Carrasco DR, Tonon G, Huang Y, Zhang Y, Sinha R, Feng B, et al. High-resolution genomic profiles define distinct clinico-pathogenetic subgroups of multiple myeloma patients. Cancer Cell 2006;9:313–25.
24. Brunet JP, Tamayo P, Golub TR, Mesirov JP. Metagenes and molecular pattern discovery using matrix factorization. Proc Natl Acad Sci U S A 2004;101:4164–9.
25. Storey JD. A direct approach to false discovery rates. J Roy Stat Soc B 2002;64:479–98.
26. Gatza ML, Silva GO, Parker JS, Fan C, Perou CM. An integrated genomics approach identifies drivers of proliferation in luminal-subtype human breast cancer. Nat Genet 2014;46:1051–9.
27. Drier Y, Sheffer M, Domany E. Pathway-based personalized analysis of cancer. Proc Natl Acad Sci U S A 2013;110:6388–93.
28. Stuart RO, Wachsman W, Berry CC, Wang-Rodriguez J, Wasserman L, Klacansky I, et al. In silico dissection of cell-type-associated patterns of gene expression in prostate cancer. Proc Natl Acad Sci U S A 2004;101:615–20.
29. Tomlins SA, Alshalalfa M, Davicioni E, Erho N, Yousefi K, Zhao S, et al. Characterization of 1577 primary prostate cancers reveals novel biological and clinicopathologic insights into molecular subtypes. Eur Urol 2015;68:555–67.
30. Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S, et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature 2012;483:603–7.
31. Aytes A, Mitrofanova A, Lefebvre C, Alvarez MJ, Castillo-Martin M, Zheng T, et al. Cross-species regulatory network analysis identifies a synergistic interaction between FOXM1 and CENPF that drives prostate cancer malignancy. Cancer Cell 2014;25:638–51.
32. Mulholland DJ, Kobayashi N, Ruscetti M, Zhi A, Tran LM, Huang J, et al. Pten loss and RAS/MAPK activation cooperate to promote EMT and metastasis initiated from prostate cancer stem/progenitor cells. Cancer Res 2012;72:1878–89.
33. Erho N, Crisan A, Vergara IA, Mitra AP, Ghadessi M, Buerki C, et al. Discovery and validation of a prostate cancer genomic classifier that predicts early metastasis following radical prostatectomy. PLoS One 2013;8:e66855.
34. Cancer Genome Atlas Research Network. Electronic address scmo, Cancer Genome Atlas Research N. The Molecular Taxonomy of Primary Prostate Cancer. Cell 2015;163:1011–25.
35. Robinson D, Van Allen EM, Wu YM, Schultz N, Lonigro RJ, Mosquera JM, et al. Integrative clinical genomics of advanced prostate cancer. Cell 2015;161:1215–28.
36. Goldstein AS, Huang J, Guo C, Garraway IP, Witte ON. Identification of a cell of origin for human prostate cancer. Science 2010;329:568–71.
37. Wang ZA, Mitrofanova A, Bergren SK, Abate-Shen C, Cardiff RD, Califano A, et al. Lineage analysis of basal epithelial cells reveals their unexpected plasticity and supports a cell-of-origin model for prostate cancer heterogeneity. Nat Cell Biol 2013;15:274–83.
38. Baird AA, Muir TC. Membrane hyperpolarization, cyclic nucleotide levels and relaxation in the guinea-pig internal anal sphincter. Br J Pharmacol 1990;100:329–35.
39. Visvader JE. Keeping abreast of the mammary epithelial hierarchy and breast tumorigenesis. Genes Dev 2009;23:2563–77.
40. Liu H, Cadaneanu RM, Lai K, Zhang B, Huo L, An DS, et al. Differential gene expression profiling of functionally and developmentally distinct human prostate epithelial populations. Prostate 2015;75:764–76.
41. Smith BA, Sokolov A, Uzunangelov V, Baertsch R, Newton Y, Graim K, et al. A basal stem cell signature identifies aggressive prostate cancer phenotypes. Proc Natl Acad Sci U S A 2015;112:E6544–52.
42. Miyamoto DT, Zheng Y, Wittner BS, Lee RJ, Zhu H, Broderick KT, et al. RNA-Seq of single prostate CTCs implicates noncanonical Wnt signaling in antiandrogen resistance. Science 2015;349:1351–6.
43. Geiss GK, Bumgarner RE, Birditt B, Dahl T, Dowidar N, Dunaway DL, et al. Direct multiplexed measurement of gene expression with color-coded probe pairs. Nat Biotechnol 2008;26:317–25.
44. Morrison T, Hurley J, Garcia J, Yoder K, Katz A, Roberts D, et al. Nanoliter high throughput quantitative PCR. Nucleic Acids Res 2006;34:e123.
45. Martin NE, Mucci LA, Loda M, Depinho RA. Prognostic determinants in prostate cancer. Cancer J 2011;17:429–37.